WARWICK

ECONOMICS

Quantile Graphical Models : Prediction and Conditional Independence with Applications to Financial Risk Management

Alexandre Belloni, Mingli Chen,

and Victor Chernozhukov

June 2016 No: 1125

Warwick Economics Research Papers

ISSN 2059-4283 (online)

ISSN 0083-7350 (print)

QUANTILE GRAPHICAL MODELS: PREDICTION AND CONDITIONAL INDEPENDENCE WITH APPLICATIONS TO FINANCIAL RISK MANAGEMENT

ALEXANDRE BELLONI*, MINGLI CHEN‡, AND VICTOR CHERNOZHUKOV†

Abstract. We propose Quantile Graphical Models (QGMs) to characterize predictive and conditional independence relationships within a set of random variables of interest. This framework is intended to quantify the dependence in non-Gaussian settings which are ubiquitous in many econometric applications. We consider two distinct QGMs. First, Condition Independence QGMs characterize conditional independence at each quantile index revealing the distributional dependence structure. Second, Predictive QGMs characterize the best linear predictor under asymmetric loss functions. Under Gaussianity these notions essentially coincide but non-Gaussian settings lead us to different models as prediction and conditional independence are fundamentally different properties. Combined the models complement the methods based on normal and nonparanormal distributions that study mean predictability and use covariance and precision matrices for conditional independence.

We also propose estimators for each QGMs. The estimators are based on high-dimension techniques including (a continuum of) ℓ_1-penalized quantile regressions and low biased equations, which allows us to handle the potentially large number of variables. We build upon recent results to obtain valid choice of the penalty parameters and rates of convergence. These results are derived without any assumptions on the separation from zero and are uniformly valid across a wide-range of models. With the additional assumptions that the coefficients are well-separated from zero, we can consistently estimate the graph associated with the dependence structure by hard thresholding the proposed estimators.

Further we show how QGM can be used to represent the tail interdependence of the variables which plays an important role in application concern with extreme events in opposition to average behavior. We show that the associated tail risk network can be used for measuring systemic risk contributions. We also apply the framework to study financial contagion and the impact of downside movement in the market on the dependence structure of assets' return. Finally, we illustrate the properties of the proposed framework through simulated examples.

Key words: High-dimensional sparse model, tail risk, conditional independence, nonlinear correlation, penalized quantile regression, systemic risk, financial contagion, downside movement

---

Date: First version: November 2012, this version July 1, 2016.

We would like to thank Don Andrews, Debopam Bhattacharya, Xu Cheng, Valentina Corradi, Victor de la Pena, Aureo de Paula, Francis Diebold, Ivan Fernandez-Val, Jiaying Gu, Roger Koenker, Arthur Lewbel, Whitney Newey, Dong Hwan Oh, David Pacini, Andrew Patton, Elisabeth Perlman, Stephen Portnoy, Shu Shen, Senay Sokullu, Sami Stouli, Aleksey Tetenov, and Yanos Zylberberg for comments and discussions. We would also like to thank the participants from the Boston University Seminar Series, 2015 Warwick Summer Workshop, 11th World Congress of the Econometric Society, and the University of Bristol Seminar Series.

*Duke University, e-mail:abn5@duke.edu.

‡University of Warwick, e-mail:M.Chen.3@warwick.ac.uk.

†Massachusetts Institute of Technology, e-mail:vchern@mit.edu.

1. Introduction

We propose Quantile Graphical Models (QGMs) to characterize and visualize the dependence structure of a set of random variables. The proposed framework allows us to understand prediction and conditional independence between these variables. Moreover, it also enable us to focus on specific parts of the distributions of these variables such as tail events. Such understanding plays an important role in applications like financial contagion and systemic risk measuring where extreme events are the main interest for regulators. Our techniques are intended to be applied in high-dimensional settings where the total number of variables (or additional conditioning variables) is large – possibly larger than the sample size.

In this work we provide an alternative route to estimate conditional independence and predictability under asymmetric loss functions that is appealing to non-Gaussian settings. In the Gaussian setting these notions essentially coincide but to pursue these questions in non-Gaussian settings different approaches are needed. Conditional independence interpretation hinges on the equivalence between conditional probabilities and conditional quantiles to characterize a random variable. Prediction performance under asymmetric loss function hinges of the solution of a quantile regression (M-estimation problem) with non-vanishing misspecification. Although we build upon the quantile regression literature ([43, 17]), we derive new results on penalized quantile regression in high dimensional settings that enables us to handle misspecification, many controls and a continuum of additional conditioning events.

Our interest lies on understanding the dependence and prediction properties among the components of a d-dimensional random vector X_V, where the set V contains the labels of the components. Quantile graphical models (QGMs) allow us to visualize dependence for each specific quantile index τ through a graph where the set of nodes V represents the components of X_V and edges represent a relation between the corresponding components. Therefore we have a graph process indexed by $\tau \in (0, 1)$. The structure represented by the τ-quantile graph represents a local relation and can be valuable in applications where the tail interdependence (high or low quantile index) is the main interest.

A motivation of predictive quantile graphical models is to allow for a possible misspecified model. This approach was first properly justified by [8] where it was shown that we recover a suitable "best approximation" for the conditional quantile function. In particular, we can guarantee good prediction properties under asymmetric loss function. Other papers also investigated the impact of misspecification in the specification of the quantile function, see [8], [42], [47] and [1]. This work is the first to accommodate non-vanishing misspecification for high-dimensional quantile regression.

Our work is complementary to a large body of work that focused on the case of jointly Gaussian random variables [45]. Indeed, under Gaussianity the "population" conditional independence graphs and prediction graphs coincide. In such setting, it is well known that conditional independence structure is completely characterized by the covariance matrix of the random variables of interest. Indeed, a zero entry in the precision matrix (inverse of the covariance matrix) identifies a pair of conditionally independent variables. Thus the precision matrix can be directly translated into a Gaussian graphical model (GGM) which can be used to study the interdependence. Further this approach characterize the conditional mean predictability of one set of the variables by linear combinations of the other variables.

The network produced by QGMs has several important features. First, it enables different strength of the links in different directions. This is important because for undirected networks, the distinction between exposure and contribution is unclear. Second, compared with Gaussian Graphical Models (which is characterized by the covariance matrix), QGMs are able to capture the tail interdependence through estimating at up or low quantiles. Third, QGMs can capture the asymmetric dependence structure at different quantiles, which can be particularly useful in applications (e.g., stock market returns, exchange rate dependence). In addition, by considering all the quantiles we can characterize the conditional independence structure between the variables. This is useful specially when the variables are not jointly Gaussian distributed, in which case the covariance matrix cannot completely characterize establish conditional independence.

We also provide estimation methods to learn QGMs from data. The estimators are geared to cover high-dimensional settings where the size of the model is large relative to the sample size. These estimators are based on ℓ_1-penalized quantile regression and low biased equations. For the CIQGMs, under mild regularities conditions, we discuss rates of convergence and properties of the selected graph structure that hold uniformly over a large class of data generating process. We provide simultaneously valid confidence regions (post-selection) for the coefficients of the CIQGM that are uniformly valid despite of possible model selection mistakes. Furthermore, based on proper thresholding, recovery of the QGMs pattern is possible when coefficients are well separated from zero which parallel the results for graph recovery in the Gaussian case based on the estimation of the precision matrix. (Similar to the graph recovery in the Gaussian case such exact recover is subject to the lack of uniformity validity critiques of Leeb and Pötscher [48].) For the PQGMs, we provide an estimate that achieves an adaptive rate of convergence which might differ for different conditioning events.

QGMs can play an important role in applications where tail events are relevant. With certain rescaling of the edge weights, we are able to capture the importance of each node or measuring its systemic risk contribution. In parallel with [5], many approaches to systemic risk measurement fit naturally into the QGMs. For example, one can view the $\Delta CoVaR_\tau^{b|a}$, $a, b \in V$ (suitably scaled), as a measure of the impact of firm a on firm b, as the weight in the edge of a QGM at quantile τ. Then, the systemic risk of firm a, $\Delta CoVaR_\tau^{sys|a}$ which measures contributions of individual firms to systemic network event, equals to the sum of coefficients over $b \in V$, $\sum_{b\in V} \Delta CoVaR_\tau^{b|a}$. Similarly, the sum over $a \in V$ measures exposures of individual firms to systemic shocks from the network.

QGMs can also be used to study contagion and network spillover effects since it is useful for studying tail risk spillovers. We consider the study of international financial contagion in volatilities, specializing in estimating the risk transmission channels, see [28] for an overview on international financial contagions. After estimating the risk transmission channels, we can use our $\Delta CoVaR$ measure to calculate the contribution and exposure of each country to the whole market. Our method applies to the case where many countries involved, overcome the problem of curse of dimensionality that traditional methods might have.

Understanding the dependence between stock returns plays a key role in hedging strategies. However, these strategies are critical precisely during downside movement of the market. The union of QGMs can be more informative in representing conditional independence than Gaussian graphical models in this setting. Indeed, recent empirical evidence [7, 6, 56] points to non-Gaussianity of the distribution of stock returns, especially during market downturns. Further, hedging decisions are typically interested on extreme outcomes rather than average outcomes. Finally, it is also instructive to understand how the dependence (and policies) would change as the downside movement of the market becomes more extreme. This application motivated us to consider conditional QGMs that extend the previous models to be conditional on additional events (e.g. downside movement of the market).

Regarding the conditional independence structure for high dimensional models, this paper relates to the large statistic literature on estimating high dimensional Gaussian Graphical Models. It is well known that recovering the structure of an undirected Gaussian graph is equivalent to recovering the support of the precision matrix, i.e. covariance matrix estimation, [30] and [45, 29, 35]. Several methods for covariance matrix estimation involves hypothesis testing, [35, 32, 33, 34]. In the high-dimensional setting, [54] propose neighborhood selection with the Lasso for each node in the graph and combine the results column-by-column to get the final Gaussian graphs. [66, 9, 39] directly estimate the precision matrix through penalizing the log-likelihood function directly. Other refinement estimators including [65, 23, 52, 59, 51]. [50] extended the result to a more general class of models called nonparanormal models or semiparametric Gaussian copula models, i.e., the variables follow a joint normal distribution after a set of unknown monotone transformations. See also [49, 63, 64]. However, all those methods assume the (transformed) random matrix follows a joint normal distribution. The proposed work provides a complementary method for additional settings by giving up efficiency in the Gaussian case.

Notation. For an integer k, we let $[k] := \{1, \ldots, k\}$ denote the set of integers from 1 to k. For a random variable X we note by $\mathcal{X}$ its support. We use the notation $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. We use $\|v\|_k$ to denote the p-norm of a vector v. We denote the ℓ_0-"norm" by $\|\cdot\|_0$ (i.e., the number of nonzero components). Given a vector $\delta \in \mathbb{R}^p$, and a set of indices $T \subset \{1, ..., d\}$, we denote by δ_T the vector in which $\delta_{Tj} = \delta_j$ if $j \in T$, $\delta_{Tj} = 0$ if $j \notin T$. The check function is denoted by $\rho_\tau(t) = t(\tau - 1\{t \leqslant 0\})$.

2. Quantile Graphical Models

In this section we describe quantile graphical models associated with a d-dimensional random vector $X = X_V$ where the set $V = [d] = \{1, \ldots, d\}$ denotes the labels of the components. These models aim to provide a description of the interdependence between the random variables in X_V. In particular, they induce graphs that allow for visualization of dependence structures. However, different models arise from different objectives as we discuss below.

2.1. Conditional Independence Quantile Graphs.

Conditional independence graphs have been used to provide a visualization and insight on the dependence structure between random variables. Each node of the graph is associated with a component of X_V. We denote the conditional independence graph as $G^I = (V, E^I)$ where G^I is an undirected graph with vertex set V and edge set E which is represented by an adjacency matrix ($E^I_{a,b} = 1$ if the edge $(a, b) \in G^I$, and $E^I_{a,b} = 0$ otherwise). An edge (a, b) is not contained in the graph if and only if

$$X_a \perp X_b \ \mid \ X_{V\setminus\{a,b\}}, \tag{2.1}$$

namely X_b and X_a are independent conditional on all remaining variables $X_{V\setminus\{a,b\}} = \{X_k; k \in V\setminus\{a, b\}\}$.

Comment 2.1 (Conditional Independence Under Gaussianity)**.** In the case that X is jointly normally distributed, $X_V \sim N_d(0, \Sigma)$ with Σ as the covariance matrix of X_V, the conditional independence structure between two components is determined by the inverse of covariance matrix, i.e. the precision matrix $\Theta = \Sigma^{-1}$. It follows that the nonzero elements in the precision matrix corresponds to the nonzero coefficients of the associated (high dimensional) mean regression. The family of Gaussian distributions with this property is known as a Gauss-Markov random field with respect to the graph G. This observation has motivated a large literature [45] and some extension that allow transformations of Gaussian variables. ■

Our main interest is to allow for non-Gaussian distributions. In order to achieve a tractable concept in such generality, we use that (2.1) occurs if and only if

$$F_{X_a}(\cdot \mid X_{V\setminus\{a\}}) = F_{X_a}(\cdot \mid X_{V\setminus\{a,b\}}) \ \text{ for all } \ X_{V\setminus\{a\}} \in \mathcal{X}_{V\setminus\{a\}}. \tag{2.2}$$

In turn, by the equivalence between conditional probabilities and conditional quantiles to characterize a random variable, we have that (2.1) occurs if and only if

$$Q_{X_a}(\tau|X_{V\setminus\{a\}}) = Q_{X_a}(\tau|X_{V\setminus\{a,b\}}) \quad \text{for all } \tau \in (0, 1), \ \text{ and } \ X_{V\setminus\{a\}} \in \mathcal{X}_{V\setminus\{a\}}. \tag{2.3}$$

For a quantile index $\tau \in (0, 1)$, we define the τ-quantile conditional independence graph as the directed graph $G(\tau) = (V, E^I(\tau))$ with vertex set V and edge set $E^I(\tau)$. An edge (a, b) is not contained in the edge set $E^I(\tau)$ if and only if

$$Q_{X_a}(\tau \mid X_{V\setminus\{a\}}) = Q_{X_a}(\tau|X_{V\setminus\{a,b\}}) \ \text{ for all } \ X_{V\setminus\{a\}} \in \mathcal{X}_{V\setminus\{a\}}. \tag{2.4}$$

By the equivalence between (2.2) and (2.3), the union of τ-quantile graphs over $\tau \in (0, 1)$ represents the conditional independence structure of X, namely $E^I = \cup_{\tau\in(0,1)}E^I(\tau)$. This motivates us to consider a relaxation of (2.1). For a set of quantile indices $\mathcal{T} \subset (0, 1)$, we say that

$$X_a \perp_{\mathcal{T}} X_b \ \mid \ X_{V\setminus\{a,b\}}, \tag{2.5}$$

X_a and X_b are $\mathcal{T}$-conditionally independent given $X_{V\setminus\{a,b\}}$, if (2.4) holds for all $\tau \in \mathcal{T}$. Thus, we have that (2.1) implies (2.5).We define the quantile graph associated with $\mathcal{T}$ as $G^I(\mathcal{T}) = (V, E^I(\mathcal{T}))$ where

$$E^I(\mathcal{T}) = \cup_{\tau\in\mathcal{T}}E^I(\tau).$$

Although the conditional independence concept relates to all quantile indices, the quantile characterization described above also lends itself to quantile specific impacts which can be of independent interest.[1]

[1]For example, we might be interested in some extreme events which typically correspond to crises in financial systems.

Comment 2.2 (Simulation and Conditional Independence)**.** The tools developed here can also be useful to develop simulation tools for high-dimensional random vectors. For example, we can simulate a random vector X as follows

$$X_1 \sim Q_{X_1}(U_1), \quad X_2 \sim Q_{X_2}(U_2 \mid X_1), \quad X_3 \sim Q_{X_3}(U_3 \mid X_1, X_2), \quad \ldots, \quad X_d \sim Q_{X_d}(U_d \mid X_1, \ldots, X_{d-1})$$

where $U_j \sim \text{Uniform}(0,1)$ and estimates of the conditional quantiles can be obtained based on a sample $(X_i \in \mathbb{R}^d)_{i=1}^n$ and the tools discussed here. It is also clear that the order of the procedure can impact the estimation. In particular, if most variables are independent of (say) X_d, skipping them from the process are likely to increase the accuracy of the simulation procedure.

2.2. **Prediction Quantile Graphs.** Prediction Quantile Graph Models (PQGMs) are concerned with prediction accuracy (instead of conditional independence as in Section 2.1). More precisely, for each $a \in V$, we are interested on the predicting X_a based on linear combinations of the remaining variables, $X_{V\setminus\{a\}}$, where accuracy is measured with respect to an asymmetric loss function. Formally, PQGMs measure accuracy as

$$\mathcal{L}_a(\tau \mid V \setminus \{a\}) = \min_{\beta} \mathrm{E}[\rho_\tau(X_a - X_{-a}'\beta)] \tag{2.6}$$

where $X_{-a} = (1, X_{V\setminus\{a\}}')'$, and we use the asymmetric loss function $\rho_\tau(t) = (\tau - 1\{t \leqslant 0\})\, t$ is the check function used in Koenker and Basset (1978).

Importantly, PQGMs are concerned with the best linear predictor under the asymmetric loss function ρ_τ. This is a fundamental distinction with respect to CIQGMs discussed in Section 2.1 where the specification of the conditional quantile was approximately a linear function of transformations Z_a. Indeed, we note that under suitable conditions the linear predictor that solves the minimization problem in (2.6) approximates the conditional quantile regression as shown in [13]. (In fact, the conditional quantile function would be linear if the vector X_V was jointly Gaussian.) However, PQGMs do not assume that the conditional quantile function of X_a is well approximated by a linear function and instead it focuses on the best linear predictor.

In principle each component of X_V can have predictive power for other components. However, we say that X_b is predictively uninformative for X_a given $X_{V\setminus\{a,b\}}$ if

$$\mathcal{L}_a(\tau \mid V \setminus \{a\}) = \mathcal{L}_a(\tau \mid V \setminus \{a,b\}) \quad \text{for all} \quad \tau \in (0,1).$$

Therefore, considering a linear function of X_b does not improve our performance of predicting X_a with respect to the asymmetric loss function ρ_τ.

Again we can visualize the prediction relations using a graph process indexed by $\tau \in (0,1)$. PQGMs allow us to visualize which variables are predictively informative to another variable by using a directed graph $G^P(\tau) = (V, E^P(\tau))$ where edge (a,b) is in the graph only if X_b is predictively informative for X_a given $X_{V\setminus\{a,b\}}$ at the quantile τ. Finally we define the PQGM associated with $\mathcal{T} \subset (0,1)$ as $G^P(\mathcal{T}) = (V, E^P(\mathcal{T}))$ where

$$E^P(\mathcal{T}) = \cup_{\tau\in\mathcal{T}} E^P(\tau).$$

2.3. **$\mathcal{W}$-Conditional Quantile Graphical Models.** In this section we discuss an useful extension of the QGMs discussed in Sections 2.1 and 2.2. It allows for conditioning on additional events $\varpi \in \mathcal{W}$ which is a possibly infinite collection of events. (By abuse of notation, we let ϖ to denote the event and also the index of such event. For example, we write $\mathrm{P}(\varpi)$ as a shorthand for $\mathrm{P}(W \in \Omega_\varpi)$) This is motivated by several applications where the interdependence between the random variables in X_V maybe substantially impacted by additional observable events. This general framework allows to accommodate different forms of conditioning. The main implication of this extension is that the QGMs are now graph processes indexed by $\tau \in \mathcal{T} \subset (0,1)$ and $\varpi \in \mathcal{W}$.

In order to generalize CIQGMs, we say that X_a and X_b are $(\mathcal{T}, \varpi)$-conditionally independent,

$$X_a \perp_{\mathcal{T}} X_b \ \mid \ X_{V\setminus\{a,b\}}, \varpi \tag{2.7}$$

provided that for all $\tau \in \mathcal{T}$ we have

$$Q_{X_a}(\tau|X_{V\setminus\{a\}}, \varpi) = Q_{X_a}(\tau|X_{V\setminus\{a,b\}}, \varpi). \tag{2.8}$$

The conditional independence edge set associated with (τ, ϖ) is defined analogously as before. We denote them by $E^I(\tau, \varpi)$ and $E^I(\mathcal{T}, \varpi) = \cup_{\tau\in\mathcal{T}} E^I(\tau, \varpi)$ for each $\varpi \in \mathcal{W}$.

The extension of PQGMs proceeds by defining the accuracy under the asymmetric loss function conditionally on ϖ. More precisely, we define

$$\mathcal{L}_a(\tau \mid V\setminus\{a\}, \varpi) = \min_\beta \mathrm{E}[\rho_\tau(X_a - X_{-a}'\beta) \mid \varpi]. \tag{2.9}$$

The predictive edge set associated with (τ, ϖ) is also defined analogously as before. We denote as $E^P(\tau, \varpi)$ and $E^P(\mathcal{T}, \varpi) = \cup_{\tau\in\mathcal{T}} E^P(\tau, \varpi)$.

Example 1 (Predictive QGMs of Stock Returns Under Downside Market Movement)**.** Hedging decisions rely on the dependence of the returns of various stocks. However, hedging's performance is more relevant during downside movements of the market. In such setting it is of interest to understand interdependence conditionally on downside movements. We can parameterize the downside movements by using a random variable M, which denotes a market index, and condition the on the event $\Omega_\varpi = \{M \leqslant \varpi\}$. This allows us to define conditional predictive quantile graphical models $G^P(\tau, \varpi) = (V, E^P(\tau, \varpi))$, for each $\varpi \in \mathcal{W}$. ■

3. Estimators for High-Dimensional Quantile Graphical Models

In this section we propose and discuss estimators for QGMs introduced in Section 2. Throughout this section it is assumed that we observe i.i.d. observations of the d-dimensional random vector X_V, namely $\{X_{iV} : i = 1, \ldots, n\}$. Based on the data, unless additional assumptions are imposed we cannot estimate the quantities of interest for all $\tau \in (0,1)$. We will consider a (compact) set of quantile index $\mathcal{T} \subset (0,1)$. Nonetheless, the estimators are intended to handle high dimensional models.

3.1. **Estimators for Conditional Independence Quantile Graphs.** We will consider a conditional quantile representation for each $a \in V$. It is based on transformations of the original covariates $X_{V\setminus\{a\}}$ that creates a p-dimensional vector $Z^a = Z^a(X_{V\setminus\{a\}}) \in \mathbb{R}^p$ so that

$$Q_{X_a}(\tau|X_{V\setminus\{a\}}) = Z^a\beta_{a\tau} + r_{a\tau}, \quad \beta_{a\tau} \in \mathbb{R}^p, \quad \text{for all } \tau \in \mathcal{T} \tag{3.10}$$

where $r_{a\tau}$ denotes a small approximation error. For $b \in V \setminus \{a\}$ we let $I_a(b) := \{j : Z^a_j \text{ depends on } X_b\}$. That is, $I_a(b)$ contains the components of Z^a that are functions of X_b. Under correct specification, if X_a and X_b are conditionally independent, we have $\beta_{a\tau,j} = 0$ for all $j \in I_a(b)$, $\tau \in (0,1)$.

This allows us to connect the conditional independence quantile graph estimation problem with a model selection within quantile regression. Indeed, the representation (3.10) has been used in several quantile regression models, see [43]. Under mild conditions this model allows us to identify the process $(\beta_{a\tau})_{\tau\in\mathcal{T}}$ as the solution of the following moment equation

$$\mathrm{E}[(\tau - 1\{X_a \leqslant Z^a\beta_{a\tau} + r_{a\tau}\})Z^a] = 0. \tag{3.11}$$

In order to allow a flexible specification, so that the approximation errors are negligible, it is attractive to consider a high-dimensional vector of Z^a where its dimension p is possibly larger than the sample size. In turn, having a large number of technical controls creates an estimation challenge if the number of coefficients p is not negligible with respect to the sample size n. In such high dimensional setting a widely applicable condition that makes estimation possible is approximate sparsity [37, 19, 20]. Formally we require

$$\max_{a\in V}\sup_{\tau\in\mathcal{T}} \|\beta_{a\tau}\|_0 \leqslant s, \quad \max_{a\in V}\sup_{\tau\in\mathcal{T}} \{\mathrm{E}[r^2_{a\tau}]\}^{1/2} \lesssim \sqrt{s/n}, \quad \text{and} \quad \max_{a\in V}\sup_{\tau\in\mathcal{T}} |\mathrm{E}[f_{a\tau}r_{a\tau}Z^a]| = o(n^{-1/2}) \tag{3.12}$$

where the sparsity parameter s of the model is allowed to grow (at a slower rate) as n grows, and $f_{a\tau} = f_{X_a|X_{V\setminus\{a\}}}(Q_{X_a}(\tau \mid X_{V\setminus\{a\}}) \mid X_{V\setminus\{a\}})$ denotes the conditional probability density function evaluated at the corresponding conditional quantile value. The sparsity has implications on the maximum degree of the associated quantile graph.

Algorithm 3.1 below contains our proposal to estimate $\beta_{a\tau}, a \in V, \tau \in \mathcal{T}$. It is based on three procedures in order to overcome the high-dimensionality. In the first step we apply a (post-)ℓ_1-penalized quantile regression. The second step applies (post-)Lasso where the data is weighted by the conditional density function at the conditional quantile. [2] Finally the third step relies on (orthogonal) score function that provides immunity to (unavoidable) model selection mistakes.

There are several parameters that need to be specified for Algorithm 3.1. The penalty parameter $\lambda_{V\mathcal{T}}$ is chosen to be larger than the ℓ_∞-norm of the (rescaled) score at the true quantile function. The work in [10] exploits the fact that this quantity is pivotal in their setting. Here additional correlation structure could impact and the distribution is pivotal only for each $a \in V$. The penalty is based on the maximum of the quantiles of the following random variables (each with pivotal distribution), for $a \in V$

$$\Lambda_{a\mathcal{T}} = \sup_{\tau\in\mathcal{T}}\max_{j\in[p]} \frac{|\mathbb{E}_n[(1\{U \leqslant \tau\} - \tau)Z^a_j]|}{\sqrt{\tau(1-\tau)}\widehat{\sigma}_{aj}} \tag{3.13}$$

[2] We note that an estimate for $f_{a\tau}$ is available from ℓ_1-penalized quantile regression estimators for $\tau + h$ and $\tau - h$ where h is a bandwidth parameter, see [43, 21].

where $\{U_i : i = 1, \ldots, n\}$ are i.i.d. uniform $(0, 1)$ random variables, and $\widehat{\sigma}^2_{aj} = \mathbb{E}_n[(Z^a_j)^2]$ for $j \in [p]$. The penalty parameter $\lambda_{V\mathcal{T}}$ is defined as

$$\lambda_{V\mathcal{T}} := \max_{a \in V} \Lambda_{a\mathcal{T}}(1 - \xi/|V| \mid X_{-a}),$$

that is, the maximum of the $1 - \xi/|V|$ conditional quantile of $\Lambda_{a\mathcal{T}}$ given in (3.13). Regarding the parameters for the weighted Lasso in Step 2, we recommend a (theoretically valid) iterative choice. We refer to the Appendix A where we collect the implementation details of the algorithm. We denote by $\|\beta\|_{1,\widehat{\sigma}} := \sum_j |\beta_j|\widehat{\sigma}_{aj}$ the standardized version of the ℓ_1-norm.

Algorithm 3.1. (Conditional Independence QGM estimator.) *For each $a \in V$, and $j \in [p]$, and $\tau \in \mathcal{T}$*

Step 1. *Compute $\widehat{\beta}_{a\tau}$ from $\|\cdot\|_{1,\widehat{\sigma}}$-penalized τ-quantile regression of X_a on Z^a with penalty $\lambda_{V\mathcal{T}}\sqrt{\tau(1-\tau)}$.*
Compute $\widetilde{\beta}_{a\tau}$ from τ-quantile regression of X_a on $\{Z^a_k : |\widehat{\beta}_{a\tau k}| \geqslant \lambda_{V\mathcal{T}}\sqrt{\tau(1-\tau)}/\widehat{\sigma}_k\}$.

Step 2. *Compute $\widetilde{\gamma}^j_{a\tau}$ from the post-Lasso estimator of $f_{a\tau}Z^a_j$ on $f_{a\tau}Z^a_{-j}$.*

Step 3. *Construct the score function $\widehat{\psi}_i(\alpha) = (\tau - 1\{X_a \leqslant Z^a_j\alpha + Z^a_{-j}\widetilde{\beta}_{a\tau}\})f_{a\tau i}(Z^a_j - Z^a_{-j}\widetilde{\gamma}^j_{a\tau})$, and for $L_{a\tau j}(\alpha) = |\mathbb{E}_n[\widehat{\psi}_i(\alpha)]|^2/\mathbb{E}_n[\widehat{\psi}^2_i(\alpha)]$, set $\check{\beta}_{a\tau,j} \in \arg\min_{\alpha \in \mathcal{A}_{a\tau j}} L_{a\tau j}(\alpha)$.*

Algorithm 1 above has been studied in [21] when it is applied to a single triple (a, τ, j). Under similar conditions, results that hold uniformly over $(a, \tau, j) \in V \times [p] \times \mathcal{T}$ are achievable (as shown in the next sections) building upon the tools developed in [10] and [25]. Algorithm 1 is tailored to achieve good rates of convergence in the ℓ_∞-norm. In particular, under standard regularity conditions, with probability going to 1 we have

$$\sup_{\tau \in \mathcal{T}} \|\beta_{a\tau} - \check{\beta}_{a\tau}\|_\infty \lesssim \sqrt{\frac{\log(p|V|n)}{n}}.$$

In order to create an estimate of $E^I(\tau) = \{(a, b) \in V \times V : \max_{j \in I_a(b)} |\beta_{a\tau,j}| > 0\}$, we define

$$\widehat{E}^I(\tau) = \left\{(a, b) \in V \times V : \max_{j \in I_a(b)} \frac{|\check{\beta}_{a\tau,j}|}{\mathrm{se}(\check{\beta}_{a\tau,j})} > \overline{\mathrm{cv}}\right\}$$

where $\mathrm{se}(\check{\beta}_{a\tau,j}) = \{\tau(1-\tau)\mathbb{E}_n[\widetilde{v}^2_{a\tau,j}]\}^{1/2}$ is an estimate of the standard deviation of the estimator, and the critical value $\overline{\mathrm{cv}}$ is set to account for the uniformity over $a \in V$, $j \in [p]$, and $\tau \in \mathcal{T}$. We discuss in the following sections a data driven procedure based on multiplier bootstrap that is theoretically valid in this high dimensional setting.

Comment 3.1 (Stepdown procedure for $\overline{\mathrm{cv}}$)**.** Setting a critical value $\overline{\mathrm{cv}}$ that accounts for the multiple hypothesis that are being tested plays an important role to select the graph $\widehat{E}^I(\tau)$. Further improvements can be obtained by considering the stepdown procedure of [58] for multiple hypothesis testing that was studied for the high-dimensional case in [24]. The procedure iteratively creates a suitable sequence of decreasing critical values. In each step only null hypotheses that were not rejected are considered to determine the critical value. Thus, as long as any hypothesis is rejected at a step, the critical value decreases and we continue to the next iteration. The procedure stops when no hypothesis in the current active set is rejected. ∎

Comment 3.2 (Estimation of conditional probability density function)**.** The algorithm above requires the conditional probability density function $f_{a\tau}$ which typically needs to be estimated in practice. It turns out that estimation of conditional quantiles yields a natural estimator for the conditional density function as

$$f_{a\tau} = \frac{1}{\partial Q_{X_a}(\tau \mid X_{-a})/\partial\tau}$$

Therefore, based on ℓ_1-penalized quantile regression estimates of $\tau + h_n$ and $\tau - h_n$ conditional quantile where $h_n \to 0$ denotes a bandwidth parameter, we have

$$\widehat{f}_{a\tau} = \frac{2h}{\widehat{Q}_{X_a}(\tau + h \mid X_{-a}) - \widehat{Q}_{X_a}(\tau - h \mid X_{-a})} \tag{3.14}$$

be an estimator of $f_{a\tau}$. Under smoothness conditions, it has the bias of order h_n^2. See [21] and the references therein for additional comments and estimators. ■

3.2. **Estimators for Prediction Quantile Graphs.** Next we discuss the specification and propose an estimator for PQGMs. In this case we are interested on studying prediction of X_a, $a \in V$, using a linear combination of $X_{V\setminus\{a\}}$ under the asymmetric loss discussed in (2.6). We will add an intercept as one of the variables for the sake of notation so that $X_{-a} = (1, X'_{V\setminus\{a\}})'$. Given the loss function ρ_τ, the target d-dimensional vector of parameters $\beta_{a\tau}$ is defined as (part of) the solution of the following optimization problem

$$\beta_{a\tau} \in \arg\min_{\beta} \; \mathrm{E}[\rho_\tau(X_a - X'_{-a}\beta)]. \tag{3.15}$$

By considering the case that d is large, the use of high-dimensional tools to achieve good estimators is of interest. The estimation procedure we propose is based on ℓ_1-penalized quantile regression. Again we consider models that satisfy an approximately sparse condition. Formally, we require the existence of sparse coefficients $\{\bar{\beta}_{a\tau} : a \in V, \tau \in \mathcal{T}\}$ such that

$$\max_{a\in V}\sup_{\tau\in\mathcal{T}} \|\bar{\beta}_{a\tau}\|_0 \leqslant s \quad \text{and} \quad \max_{a\in V}\sup_{\tau\in\mathcal{T}} \{\mathrm{E}[\{X'_{-a}(\beta_{a\tau} - \bar{\beta}_{a\tau})\}^2]\}^{1/2} \lesssim \sqrt{s/n}, \tag{3.16}$$

where (again) the sparsity parameter s of the model is allowed to grow as n grows. The high-dimensionality prevents us from using (standard) quantile regression methods and regularization methods are needed to achieve good prediction properties.

A key issue is to set the penalty parameter properly so that it bounds from above

$$\max_{a\in V}\sup_{\tau\in\mathcal{T}}\max_{j\in[d]} |\mathbb{E}_n[(1\{X_a \leqslant X'_{-a}\beta_{a\tau}\} - \tau)X_{-a,j}]|. \tag{3.17}$$

However, it is important to note that we do not assume that the conditional quantile of X_a is a linear function of X_{-a}. Under correct linear specification of the conditional quantile function, ℓ_1-penalized quantile regression estimator has been studied in [10]. The case that the conditional quantile function differs from a linear specification by a vanishing approximation errors has been considered in [40] and [21]. The analysis proposed here aims to allow for non-vanishing misspecification of the quantile function relative to a linear specification while still guarantee good rates of convergence in the ℓ_2-norm to the best linear specification. Thus we pursue the penalty parameter in the penalized quantile regression needs to account for such misspecification and is no longer pivotal as in [10].

In order to handle this issue we make a two step estimation. In the first step the penalty parameter λ_0 is conservative and set via bounds constructed based on symmetrization arguments, similar in spirit to [60, 14]. That leads to $\lambda_0 = cn^{-1/2}2(1+1/16)\sqrt{2\log(8|V|^2\{ne\}^2/\xi)}$. Although this is conservative, under mild conditions leads to estimates that can be leverage to fine tune the penalty choice. The second step uses the preliminary estimator to bootstrap (3.17) based on the tools in [24] as follows. For estimates $\widehat{\varepsilon}_i$ of the "noise" $\varepsilon_{a\tau i} = 1\{X_{ia} \leqslant X_{i,-a}'\beta_{a\tau}\} - \tau$ for $i \in [n]$, the new penalty parameter $\bar{\lambda}_{V\mathcal{T}}$ is defined as the $(1-\xi)$-quantile of

$$\Lambda := 1.1 \max_{a\in V} \sup_{\tau\in\mathcal{T}} \max_{j\in[d]\setminus\{a\}} \frac{|\mathbb{E}_n[g\widehat{\varepsilon}_{a\tau}X_j]|}{\{\mathbb{E}_n[\widehat{\varepsilon}_{a\tau}^2X_j^2]\}^{1/2}} \tag{3.18}$$

where $(g_i)_{i=1}^n$ is a sequence of i.i.d. standard Gaussian random variables. The penalty choice above adapts to the unknown correlation structure across components and quantile indices. The following algorithm states the procedure where we denote weighted ℓ_1-norms. We denote by $\|\beta\|_{1,\widehat{\sigma}} := \sum_j |\beta_j|\{\mathbb{E}_n[X_j^2]\}^{1/2}$ the standardized version of the ℓ_1-norm and $\|\beta\|_{1,\widehat{\varepsilon}} := \sum_j |\beta_j|\mathbb{E}_n[\widehat{\varepsilon}_{a\tau}^2X_j^2]\}^{1/2}$ a norm based on the estimated residuals.

Algorithm 3.2. (Predictive QGM estimator.) *For each $a \in V$, and $\tau \in \mathcal{T}$*

Step 1. *Compute $\widehat{\beta}_{a\tau}$ from $\|\cdot\|_{1,\widehat{\sigma}}$-penalized τ-quantile regression of X_a on X_{-a} with penalty λ_0.*
Compute $\widetilde{\beta}_{a\tau}$ from τ-quantile regression of X_a on $\{X_k : |\widehat{\beta}_{a\tau k}| \geqslant \lambda_0/\widehat{\sigma}_j\}$.

Step 2. *For $\widehat{\varepsilon}_{a\tau i} = 1\{X_{ia} \leqslant X_{i,-a}'\widetilde{\beta}_{a\tau}\} - \tau$ for $i \in [n]$, and $\xi = 1/n$, compute $\bar{\lambda}_{V\mathcal{T}}$ via (3.18).*

Step 3. *Compute $\widehat{\beta}_{a\tau}$ from $\|\cdot\|_{1,\widehat{\varepsilon}}$-penalized τ-quantile regression of X_a on X_{-a} with penalty $\bar{\lambda}_{V\mathcal{T}}$.*
Compute $\check{\beta}_{a\tau}$ from τ-quantile regression of X_a on $\{X_k : |\widehat{\beta}_{a\tau k}| \geqslant \bar{\lambda}_{V\mathcal{T}}/\{\mathbb{E}_n[\widehat{\varepsilon}_{a\tau}^2X_j^2]\}^{1/2}\}$.

Under regularity conditions with probability going to 1 we have

$$\max_{a\in V}\sup_{\tau\in\mathcal{T}} \|\beta_{a\tau} - \check{\beta}_{a\tau}\| \lesssim \sqrt{\frac{s\log(|V|n)}{n}}.$$

The estimate of the prediction quantile graph is given by the support of $(\check{\beta}_{a\tau})_{\tau\in\mathcal{T},a\in V}$, namely

$$\widehat{E}^P(\tau) = \left\{(a,b) \in V\times V : \ |\widehat{\beta}_{a\tau,b}| > \bar{\lambda}_{V\mathcal{T}}/\{\mathbb{E}_n[\widehat{\varepsilon}_{a\tau}^2X_b^2]\}^{1/2}\right\},$$

that is, it is induced by the covariates selected by the ℓ_1-penalized estimator. These thresholded estimators not only have the same rates of convergence as the original penalized estimator but possess additional sparsity guarantees.

3.3. **$\mathcal{W}$-Conditional Quantile Graph Models.** In order to handle the additional conditional events Ω_ϖ, $\varpi \in \mathcal{W}$, we propose to modify the Algorithms 1 and 2 based on kernel smoothing. To that extent, we assume that the observed data is of the form $\{(X_{iV}, W_i) : i = 1,\ldots,n\}$, where W_i might be defined through additional variables. Furthermore, we assume that for each conditioning event $\varpi \in \mathcal{W}$ we have access to a kernel function K_ϖ that is applied to W, to represent the relevant observations associated with ϖ (recall that we denote $\mathrm{P}(W \in \Omega_\varpi)$ as $\mathrm{P}(\varpi)$). We assume that $K_\varpi(W) = 1\{W \in \Omega_\varpi\}$.

Example 2 (Predictive QGMs of Stock Returns Under Downside Market Movements, continued)**.** In Example 1, we have W denote the market return and the conditioning event to be $\Omega_\varpi = \{W \leqslant \varpi\}$ which is parameterized by $\varpi \in \mathcal{W}$, a closed interval in $\mathbb{R}$. We might be interest on a fixed ϖ or on a family of values $\varpi \in (-\bar{\varpi}, 0]$. The latter induces $\mathcal{W} = \{\ \Omega_\varpi = \{W \leqslant \varpi\} : \varpi \in (-\bar{\varpi}, 0]\}$. The kernel function is simply $K_\varpi(t) = 1\{t \leqslant \varpi\}$.

This framework encompasses the previous framework by having $K_\varpi(W) = 1$ for all W. However, it allows for a richer class of estimands which require estimators whose properties should hold uniformly over $\varpi \in \mathcal{W}$ as well. Next we present estimators for this setting that generalize the previous methods to account for the additional conditioning on $\varpi \in \mathcal{W}$.

In what follows, we abuse the notation we use ϖ to denote not only the index but also the event Ω_ϖ. For further notational convenience, we let $u = (a, \tau, \varpi) \in \mathcal{U} = V \times \mathcal{T} \times \mathcal{W}$ so that the set $\mathcal{U}$ collects all the three relevant indices. Define the following weighted ℓ_1-norm

$$\|\beta\|_{1,\varpi} = \sum_{j\in[p]} \{\mathbb{E}_n[K_\varpi(W)(Z_j^a)^2]\}^{1/2}|\beta_j|.$$

This norm is ϖ dependent and provides the proper adjustments as we condition on different events associated with different ϖ's.

First we consider the conditional independence setting where the model is, up to small approximation errors, correctly specified. The definition of the penalty parameter will be based on the random variable

$$\Lambda_{a\mathcal{T}\mathcal{W}} = \sup_{\tau\in\mathcal{T},\varpi\in\mathcal{W},j\in[p]} \left| \frac{\mathbb{E}_n[K_\varpi(W)(1\{U \leqslant \tau\} - \tau)Z_j^a]}{\sqrt{\tau(1-\tau)}\mathbb{E}_n[K_\varpi(W)(Z_j^a)^2]^{1/2}} \right|$$

where U_i are independent uniform $(0, 1)$ random variables, and set the penalty

$$\lambda_{a\tau\varpi} = \sqrt{\tau(1-\tau)}\Lambda_{a\mathcal{T}\mathcal{W}}(1 - \gamma/\{|V|n^{1+2d_W}\} \mid X_{-a}, W)$$

where $\Lambda_{a\mathcal{T}\mathcal{W}}(1 - \xi \mid X_{-a}, W)$ is the $(1-\xi)$-quantile of Λ conditional on $\{(X_{i,-a}, W_i) : i = 1, \ldots, n\}$. Algorithm 3.3 provides the definition of the estimators.

Algorithm 3.3. ($\mathcal{W}$-Conditional Independence QGM estimator.) *For $(a, \tau, \varpi) \in V \times \mathcal{T} \times \mathcal{W}$ and $j \in [p]$*

Step 1. *Let $\widehat{\beta}_{a\tau\varpi}$ solve $\|\cdot\|_{1,\varpi}$-penalized τ-quantile regression of $K_\varpi(W)(X_a; Z^a)$ with penalty $\lambda_{a\tau\varpi}$. Compute $\widetilde{\beta}_{a\tau\varpi}$ based on the τ-quantile regression of the thresholded support of $\widehat{\beta}_{a\tau\varpi k}$, namely based on the data $K_\varpi(W)(X_a; \{Z_k^a : |\widehat{\beta}_{a\tau\varpi k}| \geqslant \lambda_{a\tau\varpi}/\mathbb{E}_n[K_\varpi(W)(Z_j^a)^2]^{1/2}\})$.*

Step 2. *Compute $\widetilde{\gamma}^j_{a\tau\varpi}$ from the post-Lasso estimator of $K_\varpi(W)f_{a\tau\varpi}Z_j^a$ on $K_\varpi(W)f_{a\tau\varpi}Z_{-j}^a$.*

Step 3. *Construct the score function*

$$\widehat{\psi}_i(\alpha) = K_\varpi(W_i)(\tau - 1\{X_{ia} \leqslant Z_{ij}^a\alpha + Z_{i,-j}^a\widetilde{\beta}_{a\tau\varpi}\})f_{a\tau\varpi i}(Z_{ij}^a - Z_{i,-j}^a\widetilde{\gamma}^j_{a\tau\varpi})$$

and for $L_{a\tau\varpi j}(\alpha) = |\mathbb{E}_n[\widehat{\psi}_i(\alpha)]|^2/\mathbb{E}_n[\widehat{\psi}_i^2(\alpha)]$, set $\check{\beta}_{a\tau\varpi,j} \in \arg\min_{\alpha\in\mathcal{A}_{a\tau\varpi j}} L_{a\tau\varpi j}(\alpha)$.

Next we consider estimators of the prediction quantile graphical models conditionally on the events in $\mathcal{W}$. Similarly to the previous case, the function K_ϖ will select the relevant subsample for each $\varpi \in \mathcal{W}$. This has implications on the penalty choices which are again not pivotal. the new penalty parameter is defined as the $(1-\xi)$-quantile of

$$\Lambda := 1.1 \max_{a\in V} \sup_{\tau\in\mathcal{T},\varpi\in\mathcal{W}} \max_{j\in[d]\backslash\{a\}} \frac{|\mathbb{E}_n[K_\varpi(W)g\widehat{\varepsilon}_{a\tau\varpi}X_j]|}{\mathbb{E}_n[K_\varpi(W)\widehat{\varepsilon}^2_{a\tau\varpi}X_j^2]^{1/2}} \tag{3.19}$$

where $(g_i)_{i=1}^n$ is a sequence of i.i.d. standard Gaussian random variables. It will also be useful to define another weighted ℓ_1-norm, $\|\beta\|_{1,\varpi\widehat{\epsilon}} := \sum_j |\beta_j|\{\mathbb{E}_n[K_\varpi(W)\widehat{\varepsilon}^2_{a\tau\varpi}X_j^2]\}^{1/2}$ the weighted. The penalty choice and weighted ℓ_1-norm adapt to the unknown correlation structure across components and quantile indices. The following algorithm states the procedure.

Algorithm 3.4. ($\mathcal{W}$-Predictive QGM estimator.) *For* $(a,\tau,\varpi) \in V \times \mathcal{T} \times \mathcal{W}$

Step 1. *Compute* $\widehat{\beta}_{a\tau\varpi}$ *from* $\|\cdot\|_{1,\varpi}$*-penalized* τ*-quantile regression of* X_a *on* X_{-a} *with penalty* $\lambda_{0\mathcal{W}}$*. Compute* $\widetilde{\beta}_{a\tau\varpi}$ *based on the* τ*-quantile regression associated with the thresholded* $\widehat{\beta}_{a\tau\varpi}$*, namely* τ*-quantile regression of* X_a *on* $\{X_k : |\widehat{\beta}_{a\tau k}| \geqslant \lambda_{0\mathcal{W}}/\{\mathbb{E}_n[K_\varpi(W)X_j^2]\}^{1/2}\}$.

Step 2. *For* $\widehat{\varepsilon}_{a\tau\varpi i} = 1\{X_{ia} \leqslant X_{i,-a}'\widetilde{\beta}_{a\tau\varpi}\} - \tau$ *for* $i \in [n]$*, and* $\xi = 1/n$*, compute* $\bar{\lambda}_{V\mathcal{T}\mathcal{W}}$ *via (3.19).*

Step 3. *Recompute* $\widehat{\beta}_{a\tau\varpi}$ *from* $\|\cdot\|_{1,\varpi\widehat{\varepsilon}}$*-penalized* τ*-quantile regression of* $K_\varpi(W)X_a$ *on* $K_\varpi(W)X_{-a}$ *with penalty* $\bar{\lambda}_{V\mathcal{T}\mathcal{W}}$*. Compute* $\check{\beta}_{a\tau\varpi}$ *from* τ*-quantile regression of* $K_\varpi(W)X_a$ *on* $\{K_\varpi(W)X_k : |\widehat{\beta}_{a\tau k}| \geqslant \bar{\lambda}_{V\mathcal{T}\mathcal{W}}/\{\mathbb{E}_n[K_\varpi(W)\widehat{\varepsilon}^2_{a\tau\varpi}X_j^2]\}^{1/2}\}$.

Comment 3.3 (Computation of Penalty Parameter over $\mathcal{W}$)**.** the penalty choices require one to maximize over $a \in V$, $\tau \in \mathcal{T}$ and $\varpi \in \mathcal{W}$. The set V is discrete and does not pose a significant challenge. However both other sets are continuous and additional care is needed. In most applications we are concerned with, $\mathcal{W}$ is a low dimensional VC class of sets and it impacts the calculation only through indicator functions. This is precisely the case of $\mathcal{T}$. It follows that only a polynomial number (in n) of different values of ϖ and τ needed to be considered. ■

4. Theoretical Results

This section is devoted to theoretical guarantees associated with the proposed estimators. We will establish rates of convergence results for the proposed estimators as well as the (uniform) validity of confidence regions. These results build upon and contribute to an increasing literature on the estimation of many process of interest under (high-dimensional) nuisance parameters.

Throughout we will provide results for the estimators based on the $\mathcal{W}$-conditional quantile graphical models as they generalized the other models by setting $K_\varpi(W) = 1$. Although some of the tools are similar, the conditional independence quantile graphical model and the predictive quantile graphical model require different estimators and are subject to different assumptions. Thus, substantial different analysis are required.

4.1. $\mathcal{W}$-Conditional Independence Quantile Graphical Model.

For $u = (a,\tau,\varpi) \in \mathcal{U}$, define the conditional τ-conditional quantile function of X_a given $X_{V\setminus\{a\}}$ and ϖ as

$$Q_{X_a}(\tau \mid X_{V\setminus a}, \varpi) = Z^a\beta_u + r_u \tag{4.20}$$

where Z^a is a p-dimensional vector of (known) transformations of $X_{V\setminus a}$, and $r(a,\tau,\varpi)$ is an approximation error. The event $\varpi \in \mathcal{W}$ will be used to further conditioning through the function $K_\varpi(W) = 1\{W \in \varpi\}$.

We let $f_{X_a|X_{V\setminus a},\varpi}(\cdot \mid X\setminus a, \varpi)$ denote the conditional density function of X_a given $X_{V\setminus a}$ and $\varpi \in \mathcal{W}$, and define $f_u := f_{X_a|X_{V\setminus a},\varpi}(Q_{X_a}(\tau \mid X_{V\setminus a},\varpi) \mid X_{V\setminus a},\varpi)$ denote the value of the conditional density function evaluated at the τ-conditional quantile. In our analysis we will consider for $u \in \mathcal{U}$

$$\underline{f}_u = \inf_{\|\delta\|=1} \frac{\mathrm{E}[f_u\{Z^a\delta\}^2 \mid \varpi]}{\mathrm{E}[\{Z^a\delta\}^2 \mid \varpi]} \text{ and } \underline{f}_{\mathcal{U}} = \min_{u\in\mathcal{U}} \underline{f}_u. \tag{4.21}$$

Moreover, for each $j \in [p]$ and $u \in \mathcal{U}$ define

$$\gamma_u^j = \arg\min_{\gamma} \mathrm{E}[f_u^2 K_\varpi(W)(Z_j^a - Z_{-j}^a\gamma)^2] \tag{4.22}$$

where $Z^a = (Z_j^a, Z_{-j}^a)$. This provides a weighted projection to construct the residuals

$$v_{uj} = f_u(Z_j^a - Z_{-j}^a\gamma_u^j)$$

that satisfy $\mathrm{E}[f_u Z_{-j}^a v_{uj} \mid \varpi] = 0$ for each $(u,j) \in \mathcal{U} \times [p]$.

The estimands of interest are $\beta_u \in \mathbb{R}^p$, $u \in \mathcal{U}$, can be written as the solution of (a continuum of) moment equations. Letting β_{uj} denote the jth component of β_u so that $\beta_{uj} \in \mathbb{R}$ solves

$$\mathrm{E}[\psi_{uj}(X, W, \beta, \eta_{uj})] = 0$$

where the function ψ_{uj} is given by

$$\psi_{uj}(X, W, \beta, \eta) = K_\varpi(W)(\tau - 1\{X_a \leqslant Z_j^a\beta + Z_{-j}^a\eta^{(1)} + \eta^{(3)}\})(Z_j^a - Z_{-j}^a\eta^{(2)})f_u$$

and the true value of the nuisance parameter is given by $\eta_{uj} = (\eta_{uj}^{(1)}, \eta_{uj}^{(2)}, \eta_{uj}^{(3)})$ with $\eta_{uj}^{(1)} = \beta_{u,-j}$, $\eta_{uj}^{(2)} = \gamma_u^j$, and $\eta_{uj}^{(3)} = r_u$. In what follows c, C denote some fixed constant, δ_n and Δ_n denote sequences going to zero with $\delta_n = n^{-\mu}$ for some sufficiently small μ.

Condition CI. *(i) Let $\mathcal{U} = V \times \mathcal{T} \times \mathcal{W}$ and $(X_i, W_i)_{i=1}^n$ denote a sequence independent and identically distributed random vectors generated accordingly to models (4.20) and (4.22) for $u \in \mathcal{U}$. Suppose that $\sup_{u\in\mathcal{U},j\in[p]}\{\|\beta_u\| + \|\gamma_u^j\|\} \leqslant C$ and $\mathcal{T}$ is a fixed compact set. The conditional distribution function of X_a given $X_{V\setminus a}$ and ϖ is absolutely continuous with continuously differentiable density $f_{X_a|X_{V\setminus a},\varpi}(t \mid X_{V\setminus a}, \varpi)$ is bounded by $\bar{f}$ and its derivative is bounded by $\bar{f}'$ uniformly over $u \in \mathcal{U}$. Further, $|f_u - f_{u'}| \leqslant L_f\|u - u'\|$ and $\|\beta_u - \beta_{u'}\| \leqslant L_\beta\|u - u'\|^\kappa$ and $\kappa \in [1/2, 1]$. The VC dimension d_W of the set $\mathcal{W}$ is fixed, and $\{Q_{X_a}(\tau \mid X_{-a}, \varpi) : (\tau, \varpi) \in \mathcal{W} \times \mathcal{T}\}$ is a VC-subgraph with VC-dimension $1 + Cd_W$ for every $a \in V$, $\mu_{\mathcal{W}} = \inf_{\varpi\in\mathcal{W}} \mathrm{P}(\varpi)$, and $\mathrm{E}[|K_\varpi(W) - K_{\varpi'}(W)|] \leqslant \bar{L}\|\varpi - \varpi'\|$. There exists $s = s_n$ such that $\sup_{u\in\mathcal{U},j\in[p]}\{\|\beta_u\|_0 + \|\bar{\gamma}_u^j\|_0\} \leqslant s$, where $\sup_{u\in\mathcal{U},j\in[p]} \|\bar{\gamma}_u^j - \gamma_u^j\| + s^{-1/2}\|\bar{\gamma}_u^j - \gamma_u^j\|_1 \leqslant C\{n^{-1}s\log(|V|pn)\}^{1/2}$. (ii) The following moment conditions hold uniformly over $u \in \mathcal{U}$ and $j \in [p]$: $\max_{a\in V}\sup_{\|\delta\|=1} \mathrm{E}[\{(X_a, Z^a)\delta\}^4 \mid \varpi] \leqslant C$, $\mathrm{E}[f_u^2(Z^a\delta)^2 \mid \varpi] \leqslant C\underline{f}_u^2\mathrm{E}[(Z^a\delta)^2 \mid \varpi]$, $c \leqslant \underline{\kappa} := \min_{a\in V}\inf_{\|\delta\|=1} \mathrm{E}[\{(X_a, Z^a)\delta\}^2 \mid \varpi]$, $\sup_{u\in\mathcal{U},\|\xi\|=1} \mathrm{E}[|(X_a, Z^a)\xi|^2 r_u^2 \mid \varpi] \leqslant C\mathrm{E}[r_u^2 \mid \varpi]$, $\max_{j\in[p],u\in\mathcal{U}} |\mathrm{E}[f_u r_u v_{uj} \mid \varpi]| \leqslant \delta_n n^{-1/2}$, $\mathrm{E}[|f_u v_{uj} Z_k^a|^2 \mid \varpi]^{1/2} \leqslant C\underline{f}_u$, $\max_{j,k}\{\mathrm{E}[|f_u v_{uj} Z_k^a|^3 \mid \varpi]^{1/3}/\mathrm{E}[|f_u v_{uj} Z_k^a|^2 \mid \varpi]^{1/2}\}\log^{1/2}(pn|V|) \leqslant \delta_n\{n\mathrm{P}(\varpi)\}^{1/6}$; (iii) Furthermore, with probability $1 - \Delta_n$: $\sup_{u\in\mathcal{U},j\in[p]} \mathbb{E}_n[r_u^2 v_{uj}^2 \mid \varpi] + \mathbb{E}_n[r_u^2 \mid \varpi] \lesssim n^{-1}s\log(p|V|n)$, $\mathrm{E}[\max_{i\leqslant n}\sup_{u\in U} |K_\varpi(W) r_{ui}|^q] \leqslant C$. (iv) For some fixed $q \geqslant 4 \vee (1 + 2d_W)$, $L_n \geqslant \mathrm{E}[\sup_{u\in\mathcal{U},j\in[p]} |v_{uj}|^q]^{1/q}$ and $M_n \geqslant \{\mathrm{E}[\|X\|_\infty^q \vee \max_{a\in V} \|Z^a\|_\infty^q]\}^{1/q}$, $diam(\mathcal{W}) \leqslant n^{1/2q}$, $n^{1/q}M_n s\sqrt{\log(pn|V|)} \leqslant \delta_n\sqrt{n\mu_{\mathcal{W}}}$, $s^3\log^3(pn|V|) \leqslant \delta_n^4 n\underline{f}_{\mathcal{U}}^2\mu_{\mathcal{W}}^3$, $s^2\log^2(p|V|n) \leqslant \delta_n^2 n\underline{f}_{\mathcal{U}}^4\mu_{\mathcal{W}}^6$, $L_n^2 n^{2/q} s\log^{3/2}(pn|V|) \leqslant \delta_n\underline{f}_{\mathcal{U}}(n\mu_{\mathcal{W}})^{1/2}$, $\{L_f + \bar{L}\}^2 M_n^2\log^2(p|V|n)/\{\mu_{\mathcal{W}}\underline{f}_{\mathcal{U}}^2\}^3 \leqslant \delta_n n$, and $n^{4/q}M_n^4\log(pn|V|)\log n \leqslant \delta_n^2 n\mu_{\mathcal{W}}^2\underline{f}_{\mathcal{U}}^2$.*

Based on Condition CI, we derive our main results regarding the proposed estimator. Moreover, we also establish new results for ℓ_1-penalized quantile regression methods that hold uniformly over the indices $u \in \mathcal{U}$. The following theorems summarize these results.

Theorem 1 (Uniform Rates of Convergence for $\mathcal{W}$-Conditional Penalized Quantile Regression)**.** *Under Condition CI, we have that with probability at least* $1 - o(1)$

$$\|\widehat{\beta}_u - \beta_u\| \lesssim \sqrt{\frac{s(1+d_W)\log(p|V|n)}{n\underline{f}_u \mathrm{P}(\varpi)}}, \quad \textit{uniformly over } u = (a,\tau,\varpi) \in \mathcal{U}$$

Moreover, the thresholded estimator $\widehat{\beta}^{\bar{\lambda}}$*, with* $\bar{\lambda} = \sqrt{(1+d_W)\log(|V|n)/n}$ *and* $\widehat{\beta}^{\bar{\lambda}}_{uj} = \widehat{\beta}_j 1\{|\widehat{\beta}_{uj}| > \bar{\lambda}\mathbb{E}_n[K_\varpi(W)(Z^a_j)^2]^{1/2}\}$*, satisfies the same rate and* $\|\widehat{\beta}^{\bar{\lambda}}\|_0 \lesssim s$*.*

Theorem 1 builds upon ideas in [10] however the proof strategy is design to derive rates that are adaptive to each $u \in \mathcal{U}$. Indeed the rates of convergence are u-dependent and they show a slower rate for rare events $\varpi \in \mathcal{W}$.

Theorem 2 (Uniform Rates of Convergence for $\mathcal{W}$-Conditional Weighted Lasso)**.** *Under Condition CI, we have that with probability at least* $1 - o(1)$

$$\|\widehat{\gamma}^j_u - \gamma^j_u\| \lesssim \frac{1}{\underline{f}_u}\sqrt{\frac{(1+d_W)s\log(p|V|n)}{n\mathrm{P}(\varpi)}} \quad \textit{and} \quad \|\widehat{\gamma}^j_u\|_0 \lesssim s, \quad \textit{uniformly over } u = (a,\tau,\varpi) \in \mathcal{U},\ j \in [p].$$

The following result establishes a uniform Bahadur representation for the final estimators.

Theorem 3 (Uniform Bahadur representation for $\mathcal{W}$-Conditional Independence QGM)**.** *Under Condition CI, the estimator* $(\check{\beta}_{uj})_{u\in\mathcal{U},j\in[p]}$ *satisfies*

$$\sigma_{uj}^{-1}\sqrt{n}(\check{\beta}_{uj} - \beta_{uj}) = \mathbb{U}_n(u,j) + O_P(\delta_n) \ \textit{in}\ \ell^\infty(\mathcal{U}\times[p]),$$

where $\sigma^2_{uj} = \tau(1-\tau)\bar{\mathrm{E}}[K_\varpi(W)v^2_{uj}]$ *and*

$$\mathbb{U}_n(u,j) := \frac{\sigma_{uj}^{-1}}{\sqrt{n}}\sum_{i=1}^n (\tau - 1\{U_i(a,\varpi) \leqslant \tau\})K_\varpi(W_i)v_{i,uj}$$

where $U_1(a,\varpi),\ldots,U_n(a,\varpi)$ *are i.i.d. uniform* $(0,1)$ *random variables, independently distributed of* $v_{1,uj},\ldots,v_{n,uj}$*.*

Theorem 3 plays a key role. However, it is important to note that the marginal distribution of $\mathbb{U}_n(u,j)$ is pivotal. Nonetheless, there is a non-trivial correlation structure between $U(a,\varpi)$ and $U(\tilde{a},\tilde{\varpi})$. In order to construct confidence regions with non-conservative guarantees, we rely on a multiplier bootstrap method. We will approximate the process $\mathcal{N} = (\mathcal{N}_{uj})_{u\in\mathcal{U},j\in[p]}$ by the Gaussian multiplier bootstrap based on estimates $\widehat{\psi}_{uj} := \widehat{\sigma}_{uj}^{-1}(\tau - 1\{X_a \leqslant Z^a\widehat{\beta}_u\})K_\varpi(W_i)\widehat{v}_{i,u}$ of $\bar{\psi}_{uj}(U,W) = \sigma_{uj}^{-1}(\tau - 1\{U(a,\varpi) \leqslant \tau\})K_\varpi(W)v_{uj}$, namely

$$\widehat{\mathcal{G}} = (\widehat{\mathcal{G}}_{uj})_{u\in\mathcal{U},j\in[p]} = \left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n g_i\widehat{\psi}_{uj}(X_i,W_i)\right\}_{u\in\mathcal{U},j\in[p]}$$

where $(g_i)_{i=1}^n$ are independent standard normal random variables which are independent from the data $(W_i)_{i=1}^n$. Based on Theorem 5.2 of [24], the following result shows that the multiplier bootstrap provides a valid approximation to the large sample probability law of $\sqrt{n}(\check{\beta}_{uj} - \beta_{uj})_{u\in\mathcal{U},j\in[p]}$ that is suitable for the construction of uniform confidence bands over the set of indices associated with $I_a(b)$ for all $a, b \in V$.

Corollary 1 (Gaussian Multiplier Bootstrap for $\mathcal{W}$-Conditional Independence Quantile Graphical Model)**.** *Under Condition CI with $\delta_n = o(\{(1+d_W)\log(p|V|n)\}^{-1/2})$, and $(1+d_W)\log(np|V|) = o(\{(n/L_n^2)^{1/7} \wedge (n^{1-2/q}/L_n^2)^{1/3}\})$, we have that*

$$\sup_{P\in\mathcal{P}_n}\sup_{t,t'\in\mathbb{R},u\in\mathcal{U},b\in V}\left|\mathrm{P}_P\left(\max_{j\in I_a(b)}\frac{|\check{\beta}_{uj}-\beta_{uj}|}{n^{-1/2}\sigma_{uj}}\in[t,t']\right) - \mathrm{P}_P\left(\max_{j\in I_a(b)}|\widehat{\mathcal{G}}_{uj}|\in[t,t'] \mid (X_i,W_i)_{i=1}^n\right)\right| = o(1)$$

4.2. ***$\mathcal{W}$*-Conditional Predictive Quantile Graph Model.** In this section we derive theoretical guarantees for the $\mathcal{W}$-condition predictive quantile estimators uniformly over $u = (a,\tau,\varpi) \in \mathcal{U}$. For each $u \in \mathcal{U}$ the estimand of interest is $\beta_u \in \mathbb{R}^p$ that correspond to the best linear predictor under asymmetric loss function, namely

$$\beta_u \in \arg\min_\beta \mathrm{E}[\rho_\tau(X_a - X_{-a}'\beta) \mid \varpi] \tag{4.23}$$

where the event $\varpi \in \mathcal{W}$ is used to further conditioning. In the analysis below the conditioning is implemented through the function $K_\varpi(W) = 1\{W \in \varpi\}$.

In the analysis of this case, the main issue is to handle the inherent misspecification of the linear form $X_{-a}'\beta_u$ with respect to the true conditional quantile. The first consequence is to handle the identification condition. Given X_{-a} and $\varpi \in \mathcal{W}$ we let $f_u := f_{X_a|X_{-a},\varpi}(X_{-a}'\beta_u \mid X_{-a},\varpi)$ denote the value of the conditional density function evaluated at $X_{-a}'\beta_u$. In our analysis we will consider for $u \in \mathcal{U}$

$$\underline{f}_u = \inf_{\|\delta\|=1}\frac{\mathrm{E}[f_u\{X_{-a}\delta\}^2 \mid \varpi]}{\mathrm{E}[(X_{-a}\delta)^2 \mid \varpi]} \quad\text{and}\quad \underline{f}_{\mathcal{U}} = \min_{u\in\mathcal{U}}\underline{f}_u. \tag{4.24}$$

We remark that $\underline{f}_u$ defined in (4.24) differs from (4.21) which is the standard conditional density at the true quantile value. It turns out that Knight's identity can be used by exploiting the first order condition associated with the optimization problem (4.23) which yields zero mean condition similar to the conditional quantile condition. A second consequence of the misspecification is the lack of pivotality of the score. Such pivotal property was convenient in the previous section to define penalty parameters and also to conduct inference. We will exploit bounds on the VC-dimension of the relevant classes of sets formally stated below.

Condition P. *(i) Let $\mathcal{U} = V \times \mathcal{T} \times \mathcal{W}$ and $(X_i, W_i)_{i=1}^n$ denote a sequence independent and identically distributed random vectors generated accordingly to models (4.23) for $u \in \mathcal{U}$. Suppose that $\sup_{u\in\mathcal{U}}\|\beta_u\| \leqslant C$ and $\mathcal{T}$ is a fixed compact set. The conditional distribution function of X_a given $X_{V\backslash a}$ and ϖ is absolutely continuous with continuously differentiable density $f_{X_a|X_{V\backslash a},\varpi}(t \mid X_{V\backslash a},\varpi)$ such that its values are bounded by $\bar{f}$ and its derivative is bounded by $\bar{f}'$ uniformly over $u \in \mathcal{U}$. Further, $|f_u - f_{u'}| \leqslant L_f\|u-u'\|$ and $\|\beta_u - \beta_{u'}\| \leqslant L_\beta\|u-u'\|^\kappa$ and $\kappa \in [1/2,1]$. The VC dimension d_W of the set $\mathcal{W}$ is fixed, and $\{1\{X_a \leqslant X_{-a}'\beta_u\} : (\tau,\varpi) \in \mathcal{W}\times\mathcal{T}\}$ is a VC-class with VC-dimension $1+d_W$ for every $a \in V$, and $\mu_{\mathcal{W}} = \inf_{\varpi\in\mathcal{W}}\mathrm{P}(\varpi)$, and $\mathrm{E}[|K_\varpi(W) - K_{\varpi'}(W)|] \leqslant \bar{L}\|\varpi-\varpi'\|$. There exists $s = s_n$ and vector such that $\sup_{u\in\mathcal{U}}\|\bar{\beta}_u\|_0 \leqslant s$, $\sup_{u\in\mathcal{U}}\|\bar{\beta}_u - \beta_u\| + s^{-1/2}\|\bar{\beta}_u - \beta_u\|_1 \leqslant \sqrt{s/n}$.*

(ii) The following moment conditions hold uniformly over $u \in \mathcal{U}$: $\max_{a\in V} \sup_{\|\delta\|=1} \mathrm{E}[\{X_{-a}'\delta\}^4 \mid \varpi] \leqslant C$, $c \leqslant \underline{\kappa} := \min_{a\in V} \inf_{\|\delta\|=1} \mathrm{E}[\{X_{-a}'\delta\}^2 \mid \varpi]$. *(iii) For some fixed $q > 4$, the sequences $M_n \geqslant \{\mathrm{E}[\max_{i\leqslant n} \|X_i\|_\infty^q]\}^{1/q}$, $q \geqslant 1 + d_W$, $diam(\mathcal{W}) \leqslant n^{1/2q}$, and the triple $(s, |V|, n)$ obey $M_n s\sqrt{\log(n|V|)} \leqslant \delta_n\sqrt{n\mu_{\mathcal{W}}}$, $s^3 \log^3(n|V|) \leqslant \delta_n n \underline{f}_{\mathcal{U}}^2 \mu_{\mathcal{W}}^2$, $\{L_f + \bar{L}\}^2 M_n^2 \log^2(p|V|n)/\{\mu_{\mathcal{W}} \underline{f}_{\mathcal{U}}^2\}^3 \leqslant \delta_n n$, $M_n^2 s \log^{3/2}(n|V|) \leqslant \delta_n \underline{f}_{\mathcal{U}} (n\mu_{\mathcal{W}})^{1/2}$ and $M_n^4 \log(n|V|)\log n \leqslant \delta_n n\mu_{\mathcal{W}}$.*

For a slack parameter $c > 1$ and confidence level $1-\xi$, the theoretical recommendation is

$$\lambda_0 = cn^{-1/2}2(1+1/16)\sqrt{2\log(8|V|^2\{ne/d_W\}^{2d_W}/\xi)}$$

Next we derive our main results regarding the proposed estimator for the best linear predictor. These results are also new ℓ_1-penalized quantile regression methods as it holds under possible misspecification of the conditional quantile function and hold uniformly over the indices $u \in \mathcal{U}$. The following theorem summarizes the result.

Theorem 4 (Uniform Rates of Convergence for $\mathcal{W}$-Conditional Penalized Quantile Regression under Misspecification)**.** *Under Condition P, we have that with probability at least $1 - o(1)$, uniformly over $u = (a, \tau, \varpi) \in \mathcal{U}$*

$$\|\widehat{\beta}_u - \beta_u\| \lesssim \sqrt{\frac{s(1+d_W)\log(|V|n)}{n\underline{f}_u \mathrm{P}(\varpi)}}$$

The data-driven choice of penalty parameter helps diminish the regularization bias and also allow (through thresholding) to obtain sparse estimators with provably rates of convergence. Moreover, the u specific penalty parameter combined with the new analysis yields an adaptive rate of convergence to each $u \in \mathcal{U}$ unlike previous works.

Comment 4.1 (Simultaneous Confidence Bands for Coefficients in PQGMs)**.** We note that in some applications we might be interested on constructing (simultaneous) confidence bands for the coefficients in PQGMs. In particular, this would include the cases practitioners are using a misspecified linear specification in a quantile regression model. Provided the conditional density function at $X_{-a}\beta_u$ can be estimated, a version of Algorithm 3.3 using the penalty parameters in Algorithm 3.4 for the initial step can deliver such confidence regions via a Multiplier bootstrap.

5. Simulations of Predictive Quantile Graph Models

In this section we perform numerical simulations to illustrate the performance of the estimators for PQGMs. We will consider several different designs. In order to compare with other proposals we will consider Gaussian and non-Gaussian examples.

5.1. **Isotropic Non-Gaussian Example.** The equivalence between a zero in the inverse covariance matrix and a pair of conditional independent variables break down for non-gaussian distribution. The nonparanormal extends Gaussian graphical models to semiparametric Gaussian copula models by transforming the variables by smooth functions. We illustrate the applicability of QGM in representing the independence structure of a set of variables when the random variables are not jointly (nonpara)normal.

Consider i.i.d. copies of an d-dimensional random vector $W = (W_1, \dots, W_d)$ from the following multivariate normal distribution, $W \sim N(0, I_{d\times d})$, where $I_{d\times d}$ is the identity matrix. Further, we generate

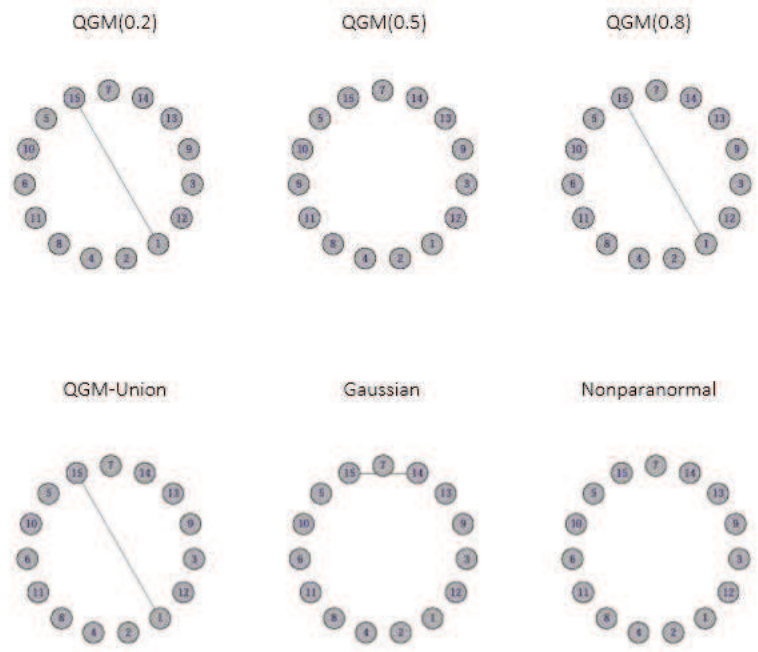
$$Y = -\sqrt{\tfrac{2}{3\pi-2}} + \sqrt{\tfrac{\pi}{3\pi-2}} W_{d-1}^2 |W_d|. \tag{5.25}$$

It follows that $E[Y] = \sqrt{\frac{\pi}{3\pi-2}}(E[|W_d|] - \sqrt{2/\pi}) = 0$ and $Var(Y) = \frac{\pi}{3\pi-2}(E[W_d^2 \cdot W_{d-1}^4] - \frac{2}{\pi}) = 1$. In addition, equation (5.25) is a location-scale-shift model in which the conditional median of the response is zero while quantile functions other than the median are nonzero. We define the vector X_V as

$$X_V = (W_1, ..., W_{d-1}, Y)'.$$

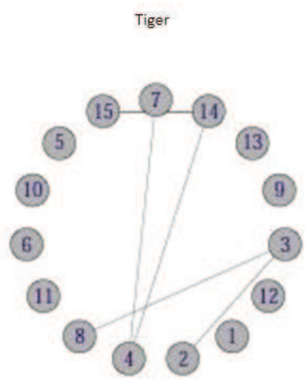
In this new set of variables, only X_{d-1} and X_d (i.e. W_{d-1} and Y) are not (conditionally) independent. Nonetheless, the new covariance matrix of X_V is still $I_{d\times d}$.

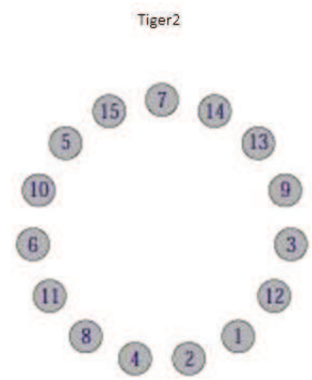
Next we consider an i.i.d. sample with a sample size of $n = 300$ and $d = 15$. We show graphs of independence structure estimated by using both the GGM and QGM(s) in this the non-Gaussian setting,



Gaussian is estimated by using graphical lasso without any transformation of X_V, and the final graph is chosen by Extended Bayesian information criterion (ebic), see [38]. Nonparanormal is estimated by using graphical lasso with nonparanormal transformation of X_V, see [50], and the final graph is chosen by ebic. Both graphs are estimated by using R-package **huge**.

We also compare our estimation results using QGM with neighborhood selection methods, e.g. TIGER of [51] in R-package **flare**, the left graph is when choosing the turning parameter to be $\sqrt{\frac{\log d}{n}}$ while the right graph is when choosing the tuning parameter to be $2\sqrt{\frac{\log d}{n}}$. Throughout, we use Tiger2 (or TIGER2) represent TIGER with penalty level $2\sqrt{\frac{\log d}{n}}$.





As expected, GGM cannot detect the correct dependence structure when the joint distribution is non-Gaussian while QGM can still represent the right independence structure.

5.2. **Gaussian Examples.** In this section we compare the numerical performance of QGM and other methods, e.g. TIGER of [51] and graphical lasso algorithm (Glasso) of [39], in recovering Gaussian Markov random field using simulated datasets. We mainly consider the Hub graph, as mentioned in [51], which also corresponds to the star network mentioned in [2, 3].

In line with [51], we generate a d-dimensional sparse graph $G = (V, E)$ represents the conditional independence structure between the variables. Let $V = \{1, ..., d\}$ correspond to variables $X = (X_1, ..., X_d)$. In our simulations, we consider 12 settings to compare these methods: (I-i) $n = 200$, $d = 10$; (I-ii) $n = 200$, $d = 20$; (I-iii) $n = 200$, $d = 40$; (I-iv) $n = 400$, $d = 10$; (I-v) $n = 400$, $d = 20$; (I-vi) $n = 400$, $d = 40$;

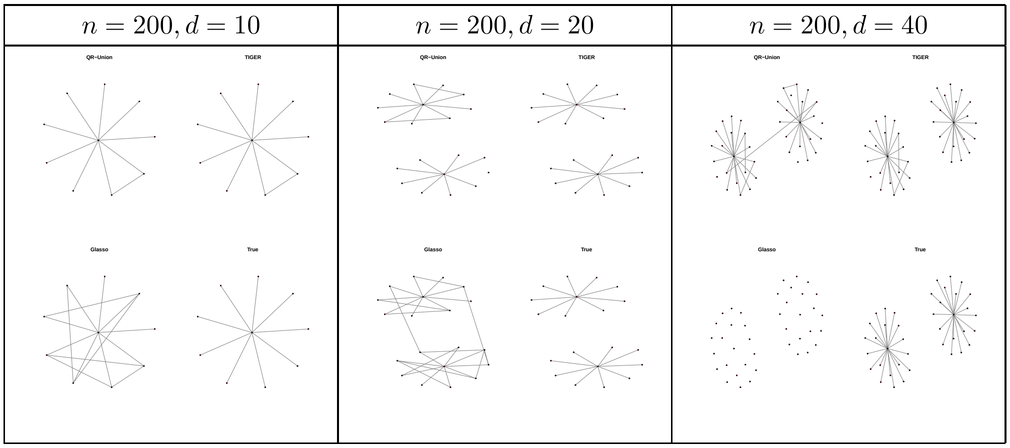
(II-i) $n = 200$, $d = 100$; (II-ii) $n = 200$, $d = 200$; (II-iii) $n = 200$, $d = 400$; (II-iv) $n = 400$, $d = 100$; (II-v) $n = 400$, $d = 200$; (II-vi) $n = 400$, $d = 400$. We adopt the following model for generating undirected graphs and precision matrices.

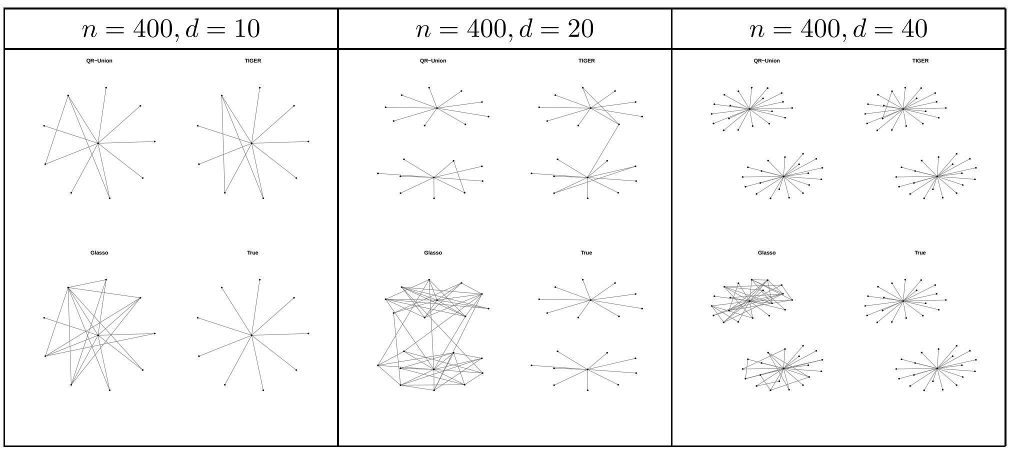
Hub graph. The d nodes are evenly partitioned into $d/20$ (or $d/10$ when $d < 20$) disjoint groups with each group contains 20 (or10) nodes. Within each group, one node is selected as the hub and we add edges between the hub and the other 19 (or 9) nodes in that group. For example, the resulting graph has 190 edges when $d = 200$ and 380 edges when $d = 400$. Once the graph is obtained, we generate an adjacency matrix $\mathbf{A}$ by setting the nonzero off-diagonal elements to be 0.3 and the diagonal elements to be 0. We calculate its smallest eigenvalue $\Lambda_{min}(\mathbf{A})$. The precision matrix is constructed as

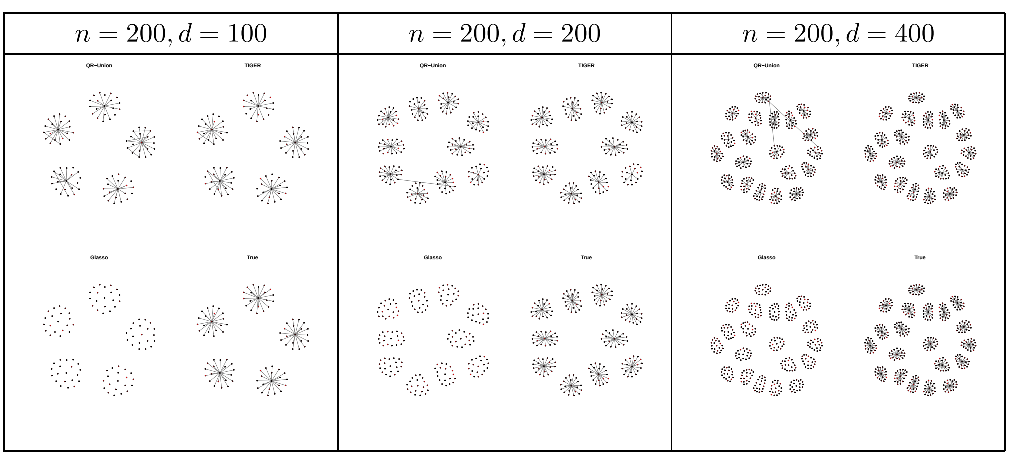
$$\Theta = \mathbf{D}[\mathbf{A} + (|\Lambda_{min}(\mathbf{A})| + 0.2) \cdot \mathbf{I}_d]\mathbf{D} \tag{5.26}$$

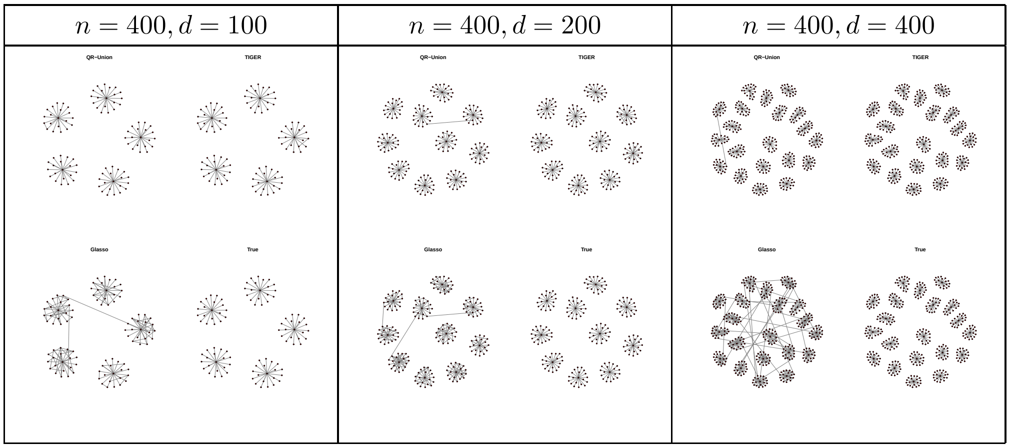
where $\mathbf{D} \in \mathbb{R}^{d \times d}$ is a diagonal matrix with $\mathbf{D}_{jj} = 1$ for $j = 1, ..., d/2$ and $\mathbf{D}_{jj} = 1.5$ for $j = d/2 + 1, ..., d$. The covariance matrix $\Sigma := \Theta^{-1}$is then computed to generate the multivariate normal data: $\mathbf{x}_1, ..., \mathbf{x}_n \sim N_d(0, \Sigma)$.

Below we provide simulation results using different estimators: QGM, TIGER and Glasso. For QGM, we use λ_I as in (3.13), and choose $c = 1.1$, $\alpha = 0.1$, $t_{thred} = \lambda/n$.


$n = 200, d = 10$
$n = 200, d = 20$
$n = 200, d = 40$
QR−Union
TIGER
Glasso
True


$n = 400, d = 10$
$n = 400, d = 20$
$n = 400, d = 40$
QR−Union
TIGER
Glasso
True


$n = 200, d = 100$
$n = 200, d = 200$
$n = 200, d = 400$
QR−Union
TIGER
Glasso
True



5.2.1. *Quantitative Comparison.* To quantify the properties of graph recovery in finite sample, we use false positive (FP) and false negative (FN) rates. For a d-dimensional graph $G = (V, E)$ with $|E| = k_s$ edges, let $\widehat{G}^{\lambda}(\mathcal{T}) = (V, \widehat{E}^{\lambda}(\mathcal{T}))$ be the estimated graph using the regularization parameter λ, the number of false positives is defined as

$$FP(\lambda) \equiv \textit{number of edges in } \widehat{E}^{\lambda} \textit{ not in } E,$$

and the number of false negatives is defined as

$$FN(\lambda) \equiv \textit{number of edges in } E \textit{ not in } \widehat{E}^{\lambda}.$$

We repeat the experiments 100 times and report the average FP and FN values with the corresponding standard deviations. We use the theoretical choice λ_I as in equation (3.13). Table 5.2.1 and 5.2.1 provide numerical comparisons of QGMs, Glasso and TIGER (with two different tuning parameters).

Table 1: Quantitative Comparison

		QGM		Glasso		Tiger1		Tiger2	
n	d	FP	FN	FP	FN	FP	FN	FP	FN
200	10	2.96	0.12	29.04	0	13.34	0	1.8	0.02
		2.11	0.47	8.29	0	4.38	0	1.76	0.2
	20	4	0.52	42.27	1.08	42.78	0	2.28	0.04
		2.96	1.00	17.38	6.17	8.28	0	2.43	0.28
	40	7.68	12.52	0	76	126.14	0.02	3.62	4.04
		3.88	4.84	0	0	15.58	0.2	3.00	3.33
400	10	2.86	0	32.2	0	13.2	0	1.96	0
		2.24	0	6.89	0	4.64	0	2.01	0
	20	3.84	0	63.17	0	42.66	0	2.82	0
		2.70	0	15.37	0	9.10	0	2.56	0
	40	6.86	0.1	66	0	125.22	0	3.72	0
		3.36	0.43	23.28	0	13.87	0	2.54	0

Table 2: Quantitative Comparison

		QGM		Glasso		Tiger1		Tiger 2	
n	d	FP	FN	FP	FN	FP	FN	FP	FN
200	100	8.54	49.02	0	198	451.7	0.14	2.74	26.84
		4.72	9.58	0	0	27.65	0.51	2.34	8.47
	200	11.48	135.54	0	398	1106.98	0.3	2.2	92.2
		4.76	14.39	0	0	332.15	0.72	2.21	16.07
	400	8	341.6	0	797.88	3201	0.9	1.96	272.32
		5.16	28.89	0	0.48	82.48	1.31	2.03	28.79
400	100	7.92	1.02	0	197.9	460.3	0	3.2	0.04
		3.92	1.46	0	0.44	32.72	0	2.81	0.28
	200	8.8	4.22	0	397.74	1237.46	0	3.18	0.3
		4.64	3.05	0	0.68	50.57	0	2.53	0.77
	400	5.8	17	0	797.8	3334.38	0	3.5	2.18
		4.05	7.01	0	0.603	86.17	0	2.55	2.38

It is clear from the tables that, in most simulation cases, QGM achieves significantly smaller errors than Glasso even if the true distribution of the data is exactly multivariate Gaussian. QGM also achieves performance comparable to TIGER, though the relative performance depends on the choosing of the tuning parameter of TIGER and tradeoff between the number of FPs and FNs. Note here all the edges are counted twice as a directed graph. So when FP=8, the actual extra linkages would be just 4.

In simulation, Tiger2 is not stable − in some simulation experiments, due to the very high penalty level, Tiger2 would not produce a graph at all.

Though in reporting the tables we choose $c = 1.1, c_{thred} = 1, \alpha = 0.1$ for both $n = 200$ and $n = 400$, in practice, when $n = 400$, choosing $c_\lambda = 1.2$, $\alpha = 0.05$ will have better graph recovery results. The intuition is, when $n = 200$, quantile regression need less penalty due to the less observations/information. All the results clearly show the performance of both QGM and TIGER improves with sample size, while the performance of Glasso is not good in general.

6. Empirical Applications of QGM

6.1. **Financial Contagion.** In this section we apply QGM for the study of international financial contagion. We focus on examining financial contagion through the volatility spillover perspective. [36] reported that international stock markets are related through their volatilities instead of returns. [31] studied the return and volatility spillovers of 19 countries and found differences in return and volatility spillovers. For a survey of financial contagion see [28]. We also illustrate how QGM can highlight asymmetric dependence between the random variables.

We use daily equity index returns, September 2009 to September 2013 (1044 observations), from Morgan Stanley Capital International (MSCI). The returns are all translated into dollar-equivalents as of September 6th 2013. We use absolute returns as a proxy for volatility. We have a total of 45 countries in our sample, there are 21 developed markets (Australia, Austria, Belgium, Canada, Denmark, France, Germany, Hong Kong, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, the United Kingdom, the United States), 21 emerging markets (Brazil, Chile, Mexico, Greece, Israel, China, Colombia, Czech Republic, Egypt, Hungary, India, Indonesia, Korea, Malaysia, Peru, Philippines, Poland, Russia, Taiwan, Thailand, Turkey), and 3 frontier markets (Argentina, Morocco, Jordan).

Below we provide a full-sample analysis of global volatility spillovers at different tails. We denote 20% quantile as Low Tail, 50% quantile as Median, 80% quantile as Up Tail. Both QGMs and GGM are estimated. Our purpose is to show the usefulness of QGM in representing nonlinear tail interdependence allowing for heteroscedasticity and to show that QGM measures correlation asymmetry by looking at behavior in the tails of the distribution (not specific to any model).

There are significant differences in the network structure in terms of volatility spillovers when using QGM and Gaussian graph. QGM permits conditional asymmetries in correlation dynamics, suited to investigate the presence of asymmetric responses. We find significant increase at the up tail interdependence between the volatility series, i.e. we find downside correlation (high volatility) are much larger than upside correlation (low volatility). This confirms findings in finance literature that financial markets become more interdependent during high volatility periods.

We also find if two countries are located in the same geographic region, with many similarities in terms of market structure and history, they tend to be closely connected (the homophily effect as in network terminology); while two economies located in separate geographic regions are less likely directly connected. We find among European Union member countries, Germany appears to play a major role in the transmission of shocks to others. While in Asia, Hong Kong, Thailand, and Singapore appears to play a major role. Among all the north and south American countries, Canada and US play a major role in risk transmission.

We also report $net\text{-}\Delta CoVaR$ to measure spillover accounting for the network (see Appendix B) for the volatility series through QGM at up tail in Figure 6.1.

Figure 6.1 shows that, globally, total volatility spillovers from Germany, France, US and Hong Kong to the others are much larger than total volatility spillovers from the others to them; while the opposite happens to Greece and Spain. Both Greece and Spain receive larger volatility spillovers from others than contribute to the others. The estimated network structure is important here as it demonstrates that shocks originating in some stock markets, e.g. Germany and Hong Kong, may be amplified in their transmission throughout the system, posing greater risks to the whole market than other shock's origination.

6.2. **Stock Returns Conditional on Market Downside Movement.** Stock markets are in general non-Gaussian. [6] find correlation asymmetries in the data and reject the null hypothesis of multivariate normal distributions at daily, weekly, and monthly frequencies, conditional on market "downside" movements. See also [53, 57] among other studies in the empirical finance literature for the non-Gaussian

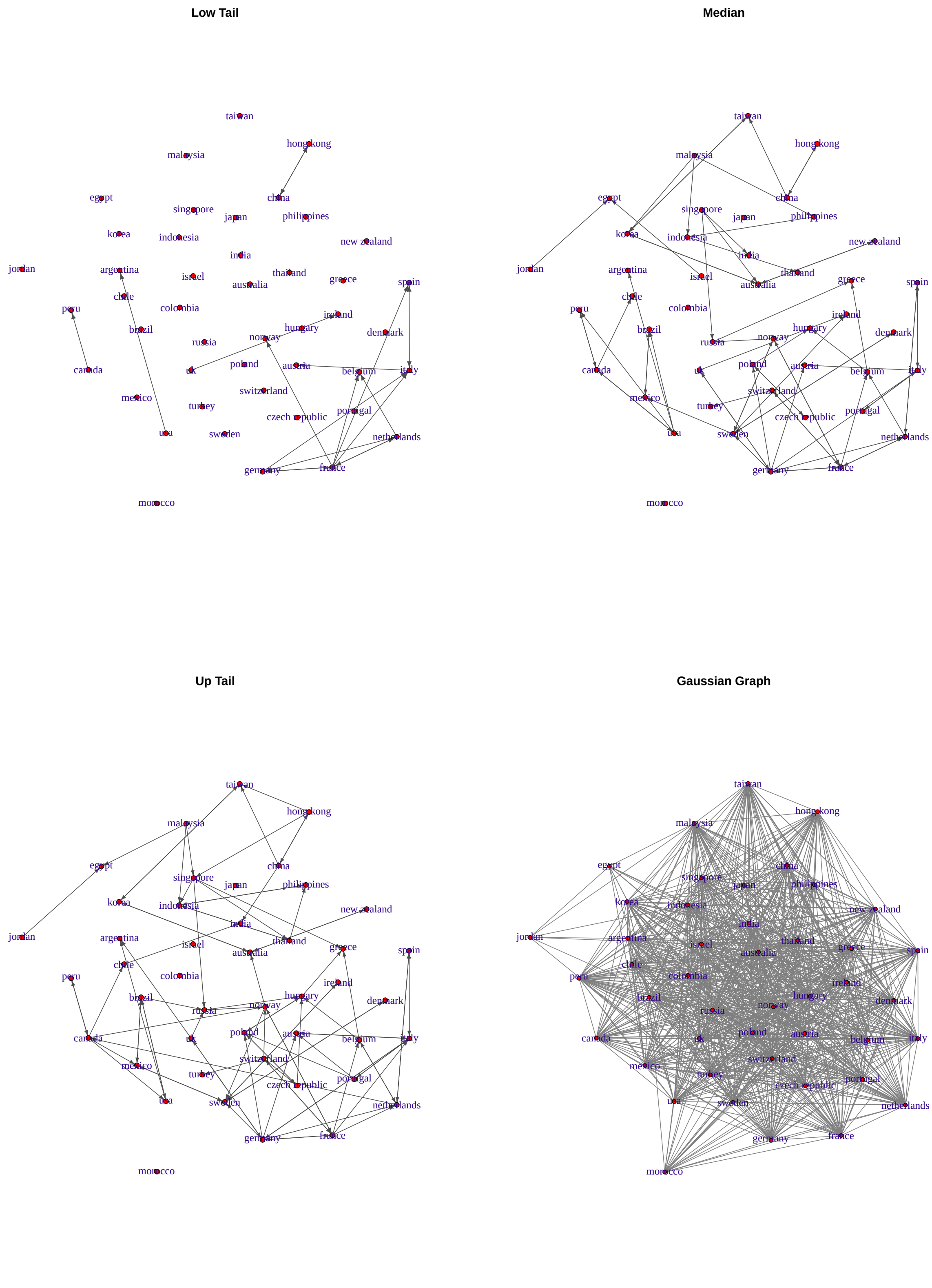


Figure 1. International Financial Contagion. Notes: We show the volatility transmission channel is asymmetric at different tails.

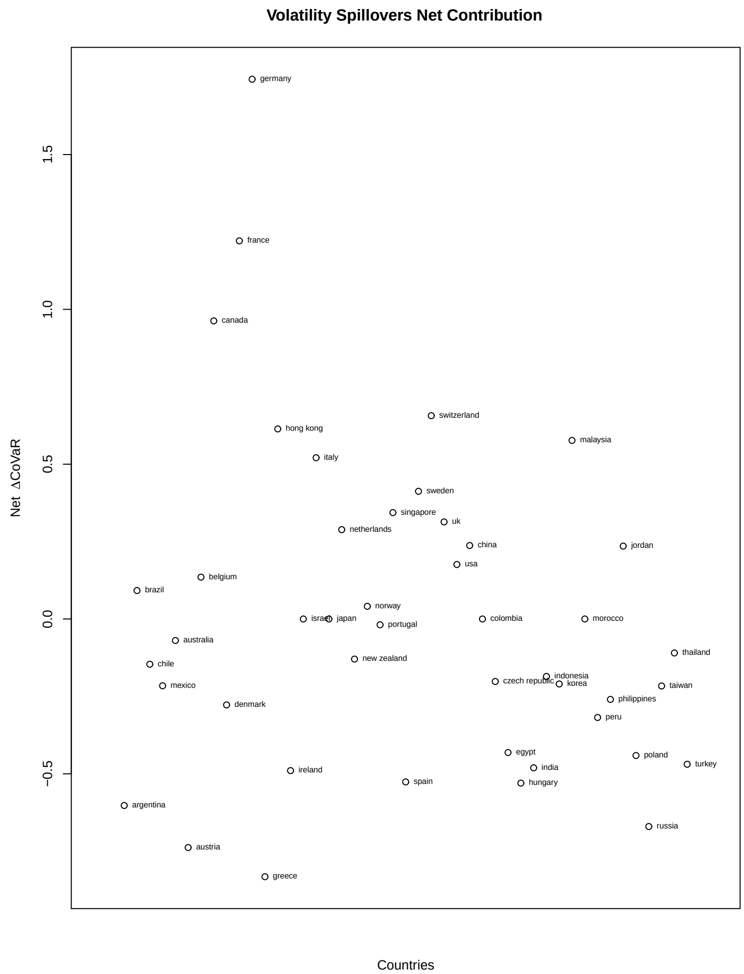


Figure 2. Net-$\Delta CoVaR$ of Each Country

feature of financial markets. Hence, generally in the financial market context, conditional correlation only conveys partial and often misleading information on the actual underlying conditional dependencies.

We contribute to the literature by showing the union of a set of QGMs can be used to obtain a conditional independence graph when the main interest lies in estimating the conditional independence structure of stocks under a market downturn. While the joint distribution of stocks considered is generally non-Gaussian, since QGM does not impose any parametric assumption on the joint distribution of stocks, the union of QGMs allows for both Gaussian and non-Gaussian joint distributions in estimating the conditional independence structure.

This will be modelled with a conditional quantile graph models. We consider the conditioning events to be $Z = \{\text{Market return} \leqslant m_u\}$ for we set $m_u = u$-th quantile of the market index return to capture downside movement of the market (note that $u = 1$ corresponds to regular market). We obtain daily stock returns from CRSP. The full sample consists of 2769 observations of daily stock returns for 86 stocks in the S&P 500 from Jan 2, 2003 to December 31, 2013. The total number of stocks is 86 due to data availability at CRSP. We define market downside as when the market index returns are below a pre-specified level and we use S&P 500 as market index. In this case, the conditioning on a particular Z corresponds simply to consider the subsample based on whether the corresponding date's market return is less equal to the u-th quantile of the market index returns. We reported the number of edges, there is no linkage between two stocks if there are conditional independent, at different subsamples in Table 6.2 below,

Table 1
Edges of Produced by Different Graph Estimators

Quantile of market index (u)	PQGM	Glasso(eBIC)	TGalasso	TIGER
0.15	406	1752	1804	3372
0.5	744	2152	2278	5734
0.75	842	2380	2478	6180
0.9	978	2461	2564	6344
1	1062	2518	2660	6290

For estimators based on QGM and GGM, the number of edges increases with the quantile index. However, potentially due to asymmetry in relations, there are significant differences between the results of QGMs and GGM. There are significantly higher interdependence in GGM. Nonetheless, increase in conditional correlation could be a result of assuming conditional normality for the return distribution – estimation bias in correlation conditional on market upside or downside moves will cause false correlation. These empirical findings support evidence from the empirical finance literature.

References

[1] Alberto Abadie, Guido W Imbens, and Fanyin Zheng. Inference for misspecified models with fixed regressors. *Journal of the American Statistical Association*, 109(508):1601–1614, 2014.

[2] Daron Acemoglu, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. Cascades in networks and aggregate volatility. Technical report, National Bureau of Economic Research, 2010.

[3] Daron Acemoglu, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. Systemic risk and stability in financial networks. Technical report, National Bureau of Economic Research, 2013.

[4] Tobias Adrian and Markus K Brunnermeier. Covar. Technical report, National Bureau of Economic Research, 2011.

[5] Torben Andersen, Tim Bollerslev, Peter Christoffersen, and Francis Diebold. Financial risk measurement for financial risk management. *Handbook of the Economics of Finance*, 2:1127–1220, 2013.

[6] Andrew Ang and Joseph Chen. Asymmetric correlations of equity portfolios. *Journal of Financial Economics*, 63(3):443–494, 2002.

[7] Andrew Ang, Joseph Chen, and Yuhang Xing. Downside risk. *Review of Financial Studies*, 19(4):1191–1239, 2006.

[8] Joshua Angrist, Victor Chernozhukov, and Iván Fernández-Val. Quantile regression under misspecification, with an application to the us wage structure. *Econometrica*, 74(2):539–563, 2006.

[9] Onureena Banerjee, Laurent El Ghaoui, and Alexandre d'Aspremont. Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *The Journal of Machine Learning Research*, 9:485–516, 2008.

[10] A. Belloni and V. Chernozhukov. ℓ_1-penalized quantile regression for high dimensional sparse models. *Annals of Statistics*, 39(1):82–130, 2011.

[11] A. Belloni and V. Chernozhukov. Least squares after model selection in high-dimensional sparse models. *Bernoulli*, 19(2):521–547, 2013. ArXiv, 2009.

[12] A. Belloni, V. Chernozhukov, D. Chetverikov, and Ying Wei. Uniformly valid post-regularization confidence regions for many functional parameters in z-estimation framework. *arXiv:1512.07619*, 2015.

[13] A. Belloni, V. Chernozhukov, and I. Fernández-Val. Conditional quantile processes based on series or many regressors. *arXiv*, 2011.

[14] A. Belloni, V. Chernozhukov, and C. Hansen. Inference methods for high-dimensional sparse econometric models. *Advances in Economics and Econometrics, 10th World Congress of Econometric Society*, III:245–295, 2013.

[15] A. Belloni, V. Chernozhukov, and C. Hansen. Inference on treatment effects after selection amongst high-dimensional controls. *Rev. Econ. Stud.*, 81:608–650, 2014.

[16] A. Belloni, V. Chernozhukov, C. Hansen, and I. Fernandez-Val. Program evaluation and causal inference with high-dimensional data. *forthcoming Econometrica*, 2016.

[17] A. Belloni, V. Chernozhukov, and K. Kato. Robust inference in high-dimensional approximately sparse quantile regression models. *arXiv:1312.7186*, 2013.

[18] A. Belloni, V. Chernozhukov, and L. Wang. Square-root-lasso: Pivotal recovery of sparse signals via conic programming. *Biometrika*, 98(4):791–806, 2011. Arxiv, 2010.

[19] Alexandre Belloni, Daniel Chen, Victor Chernozhukov, and Christian Hansen. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429, 2012.

[20] Alexandre Belloni, Victor Chernozhukov, and Christian Hansen. Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2):608–650, 2014.

[21] Alexandre Belloni, Victor Chernozhukov, and Kengo Kato. Robust inference in high-dimensional approximately sparse quantile regression models. *arXiv preprint arXiv:1312.7186*, 2013.

[22] P. J. Bickel, Y. Ritov, and A. B. Tsybakov. Simultaneous analysis of Lasso and Dantzig selector. *Annals of Statistics*, 37(4):1705–1732, 2009.

[23] Tony Cai, Weidong Liu, and Xi Luo. A constrained $\ell 1$ minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494), 2011.

[24] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41(6):2786–2819, 2013.

[25] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Gaussian approximation of suprema of empirical processes. *The Annals of Statistics*, 42(4):1564–1597, 2014.

[26] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Comparison and anti-concentration bounds for maxima of gaussian random vectors. *Probability Theory and Related Fields*, 162:47–70, 2015.

[27] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Empirical and multiplier bootstraps for supreme of empirical processes of increasing complexity, and related gaussian couplings. *arXiv preprint*, 2015.

[28] Stijn Claessens and Kristin Forbes. *International Financial Contagion*. Springer, 2001.

[29] David Roxbee Cox and Nanny Wermuth. *Multivariate dependencies: Models, analysis and interpretation*, volume 67. CRC Press, 1996.

[30] Arthur P Dempster. Covariance selection. *Biometrics*, pages 157–175, 1972.

[31] Francis X Diebold and Kamil Yilmaz. Measuring financial asset return and volatility spillovers, with application to global equity markets*. *The Economic Journal*, 119(534):158–171, 2009.

[32] Mathias Drton and Michael D Perlman. Model selection for gaussian concentration graphs. *Biometrika*, 91(3):591–602, 2004.

[33] Mathias Drton and Michael D Perlman. Multiple testing and error control in gaussian graphical model selection. *Statistical Science*, 22(3):430–449, 2007.

[34] Mathias Drton and Michael D Perlman. A sinful approach to gaussian graphical model selection. *Journal of Statistical Planning and Inference*, 138(4):1179–1200, 2008.

[35] David Edwards. *Introduction to graphical modelling*. Springer, 2000.

[36] Robert F Engle and Raul Susmel. Common volatility in international equity markets. *Journal of Business & Economic Statistics*, 11(2):167–176, 1993.

[37] Jianqing Fan, Jinchi Lv, and Lei Qi. Sparse high dimensional models in economics. *Annual review of economics*, 3:291, 2011.

[38] Rina Foygel and Mathias Drton. Extended bayesian information criteria for gaussian graphical models. In *Advances in Neural Information Processing Systems*, pages 604–612, 2010.

[39] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

[40] K. Kato. Group lasso for high dimensional sparse quantile regression models. Preprint, ArXiv, 2011.

[41] K. Knight. Limiting distributions for L_1 regression estimators under general conditions. *The Annals of Statistics*, 26:755–770, 1998.

[42] Keith Knight. Asymptotics of the regression quantile basic solution under misspecification. *Applications of Mathematics*, 53(3):223–234, 2008.

[43] R. Koenker. *Quantile regression*. Cambridge University Press, New York, 2005.

[44] Eric D Kolaczyk. *Statistical analysis of network data*. Springer, 2009.

[45] Steffen L Lauritzen. *Graphical models*. Oxford University Press, 1996.

[46] M. Ledoux and M. Talagrand. *Probability in Banach Spaces (Isoperimetry and processes)*. Ergebnisse der Mathematik undihrer Grenzgebiete, Springer-Verlag, 1991.

[47] Ying-Ying Lee. Efficiency bounds for semiparametric estimation of quantile regression under misspecification. *University of Wisconsin-Madison*, 2009.

[48] Hannes Leeb and Benedikt M Pötscher. Can one estimate the unconditional distribution of post-model-selection estimators? *Econometric Theory*, 24(02):338–376, 2008.

[49] Han Liu, Fang Han, Ming Yuan, John Lafferty, and Larry Wasserman. High-dimensional semiparametric gaussian copula graphical models. *The Annals of Statistics*, 40(4):2293–2326, 2012.

[50] Han Liu, John Lafferty, and Larry Wasserman. The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *The Journal of Machine Learning Research*, 10:2295–2328, 2009.

[51] Han Liu and Lie Wang. Tiger: A tuning-insensitive approach for optimally estimating gaussian graphical models. *arXiv preprint arXiv:1209.2437*, 2012.

[52] Weidong Liu and Xi Luo. High-dimensional sparse precision matrix estimation via sparse column inverse operator. *arXiv preprint arXiv:1203.3896*, 2012.

[53] Francois Longin and Bruno Solnik. Extreme correlation of international equity markets. *The Journal of Finance*, 56(2):649–676, 2001.

[54] Nicolai Meinshausen and Peter Bühlmann. High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, pages 1436–1462, 2006.

[55] Sahand N Negahban, Pradeep Ravikumar, Martin J Wainwright, and Bin Yu. A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers. *Statistical Science*, 27(4):538–557, 2012.

[56] Andrew J Patton. On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Econometrics*, 2(1):130–168, 2004.

[57] Andrew J Patton. Modelling asymmetric exchange rate dependence*. *International economic review*, 47(2):527–556, 2006.

[58] Joseph P Romano and Michael Wolf. Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469):94–108, 2005.

[59] Tingni Sun and Cun-Hui Zhang. Sparse matrix inversion with scaled lasso. *arXiv preprint arXiv:1202.2723*, 2012.

[60] S. A. van de Geer. High-dimensional generalized linear models and the lasso. *Annals of Statistics*, 36(2):614–645, 2008.

[61] A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes*. Springer Series in Statistics, 1996.

[62] Aad W. van der Vaart and Jon A. Wellner. Empirical process indexed by estimated functions. *IMS Lecture Notes-Monograph Series*, 55:234–252, 2007.

[63] Lingzhou Xue and Hui Zou. Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *The Annals of Statistics*, 40(5):2541–2571, 2012.

[64] Lingzhou Xue, Hui Zou, and Tianxi Cai. Nonconcave penalized composite conditional likelihood estimation of sparse ising models. *The Annals of Statistics*, 40(3):1403–1429, 2012.

[65] Ming Yuan. High dimensional inverse covariance matrix estimation via linear programming. *The Journal of Machine Learning Research*, 99:2261–2286, 2010.

[66] Ming Yuan and Yi Lin. Model selection and estimation in the gaussian graphical model. *Biometrika*, 94(1):19–35, 2007.

Appendix A. Implementation of Algorithms

Regarding the Lasso estimator, the choice of penalty level $\lambda := 1.1n^{-1/2}2\Phi^{-1}(1-\xi/N_n)$ and penalty loading $\widehat{\Gamma}_\tau = \mathrm{diag}[\widehat{\Gamma}_{\tau kk}, k \in [p]\backslash\{j\}]$ is a diagonal matrix defined by the following procedure: (1) Compute the Post Lasso estimator $\widetilde{\gamma}^j_{a\tau}$ based on λ and initial values $\widehat{\Gamma}_{\tau kk} = \max_{i\leqslant n}\|f_{ia\tau}Z^a_i\|_\infty\{\mathbb{E}_n[|f_{ia\tau}Z^a_k|^2]\}^{1/2}$. (2) Compute the residuals $\widehat{v}_i = f_{ia\tau}(Z^a_{ij} - Z^a_{i,-j}\widetilde{\gamma}^j_{a\tau})$ and update the loadings

$$\widehat{\Gamma}_{\tau kk} = \sqrt{\mathbb{E}_n[f^2_{a\tau}|Z^a_k\widehat{v}|^2]},\ k \in [p] \setminus \{j\} \tag{A.27}$$

and used them to recompute the post-Lasso estimator $\widetilde{\gamma}^j_{a\tau}$. Finally, Step 3 uses $\mathcal{A}_{a\tau j} = \{\alpha \in \mathbb{R} : |\alpha - \widetilde{\beta}_{a\tau j}| \leqslant 10\widehat{\sigma}_{aj}/\log n\}$. In the case of Algorithm 3.1 we can take $N_n = |V|p^3n^3$, in the case of Algorithm 3.3 we take $N_n = |V|p^2\{pn^3\}^{1+d_W}$.

Detailed version of Algorithm 3.1 (Conditional Independence Quantile Graphical Model)

For each $a \in V$, and $j \in [p]$, and $\tau \in \mathcal{T}$, perform the following:

(1) Run Post-ℓ_1-quantile regression of X_a on Z^a; keep fitted value $Z^a_{-j}\widetilde{\beta}_{a\tau}$, where $\widehat{\sigma}^2_j = \mathbb{E}_n[(Z^a_j)^2]$

$$\begin{aligned}&\widehat{\beta}_{a\tau} \in \arg\min_\beta \mathbb{E}_n[\rho_\tau(X_a - Z^a\beta)] + \lambda_{V\mathcal{T}}\sqrt{\tau(1-\tau)}\textstyle\sum_{j=1}^p \widehat{\sigma}_j|\beta|\\ &\widetilde{\beta}_{a\tau} \in \arg\min_\beta \mathbb{E}_n[\rho_\tau(X_a - Z^a\beta)]\ :\ \beta_j = 0\ \text{ if } |\widehat{\beta}_{a\tau j}|\widehat{\sigma}_j \leqslant \lambda_{V\mathcal{T}}\sqrt{\tau(1-\tau)}.\end{aligned}$$

(2) Run Post-Lasso of $f_{a\tau}Z^a_j$ on $f_{a\tau}Z^a_{-j}$; keep the residual $\widetilde{v}_i := f_{ia\tau}\{Z^a_{ij} - Z^a_{i,-j}\widetilde{\gamma}^j_{a\tau}\}$,

$$\begin{aligned}&\widehat{\gamma}^j_{a\tau} \in \arg\min_\gamma \mathbb{E}_n[f^2_{a\tau}(Z^a_j - Z^a_{-j}\gamma)^2] + \lambda\|\widehat{\Gamma}_\tau\gamma\|_1\\ &\widetilde{\gamma}^j_{a\tau} \in \arg\min_\gamma \mathbb{E}_n[f^2_{a\tau}(Z^a_j - Z^a_{-j}\gamma)^2]\ :\ \text{support}(\gamma) \subseteq \text{support}(\widehat{\gamma}^j_{a\tau}).\end{aligned}$$

(3) Run Instrumental Quantile Regression of $X_a - Z^a_{-j}\widetilde{\beta}_{a\tau}$ on Z^a_j using $\widetilde{v}$ as the instrument for Z^a_j,

$$\check{\beta}_{a\tau,j} \in \arg\min_{\alpha\in\mathcal{A}_\tau} \frac{\{\mathbb{E}_n[(1\{X_a \leqslant Z^a_j\alpha + Z^a_{-j}\widetilde{\beta}_{a\tau,-j}\} - \tau)\widetilde{v}]\}^2}{\mathbb{E}_n[(1\{X_a \leqslant Z^a_j\alpha + Z^a_{-j}\widetilde{\beta}_{a\tau,-j}\} - \tau)\widetilde{v}^2]}.$$

Appendix B. Incorporating network structure: CoVaR, network spillover effects, and systemic risk

Traditional risk measures, such as Value of Risk (VaR), focus on the loss of an individual institution only. CoVaR proposed by [4] measures the VaR of the whole financial system or a particular financial institution by conditioning on another institution being in distress. Thus, it relates systemic risk to tail spillover effects from individual institutions to the whole system. [4] define firm b's CoVaR at level τ conditional on a particular outcome from firm a, as the value of $CoVaR^{b|a}_\tau$ that solves

$$Pr(X_b \leqslant CoVaR^{b|a}_\tau|\mathbb{C}(X_a)) \quad = \quad \tau,$$

Detailed version of Algorithm 3.2 (Predictive Quantile Graph Model)

For each $a \in V$, and $\tau \in \mathcal{T}$, perform the following:

(1) Run ℓ_1-quantile regression of X_a on X_{-a} with penalty λ_0

$$\widehat{\beta}_{a\tau} \in \arg\min_\beta \mathbb{E}_n[\rho_\tau(X_a - X'_{-a}\beta)] + \lambda_0 \textstyle\sum_{j\in[d]\setminus\{a\}} \widehat{\sigma}_j|\beta_j|$$

where $\widehat{\sigma}_j = \{\mathbb{E}_n[X_j^2]\}^{1/2}$.

(2) Set $\widehat{\varepsilon}_{ia\tau} = 1\{X_{ia} \leqslant X'_{i,-a}\widehat{\beta}_{a\tau}\} - \tau$ for $i \in [n]$, $a \in V$ and $\tau \in \mathcal{T}$. Compute the penalty level $\lambda_{V\mathcal{T}}$ as the $(1-\xi)$-quantile conditional on the data of the random variable

$$\Lambda := 1.1 \max_{a\in V} \sup_{\tau\in\mathcal{T}} \max_{j\in[d]\setminus\{a\}} \frac{|\mathbb{E}_n[g\widehat{\varepsilon}_{a\tau}X_j]|}{\{\mathbb{E}_n[\varepsilon^2_{a\tau}X_j^2\}^{1/2}}$$

where $\{g_i : i = 1, \ldots, n\}$ is a sequence of i.i.d. standard Gaussian random variables.

(3) Run ℓ_1-quantile regression of X_a on X_{-a} with penalty $\lambda_{V\mathcal{T}}$

$$\check{\beta}^a(\tau) \in \arg\min_\beta \mathbb{E}_n[\rho_\tau(X_a - X'_{-a}\beta)] + \lambda_{V\mathcal{T}} \textstyle\sum_{j\in[d]\setminus\{a\}} |\beta_j|\{\mathbb{E}_n[\varepsilon^2_{a\tau}X_j^2\}^{1/2}$$

Detailed version of Algorithm 3.3 ($\mathcal{W}$-Conditional Independence Quantile Graphical Model)

For each $u = (a, \tau, \varpi) \in \mathcal{U} = V \times \mathcal{T} \times \mathcal{W}$, and $j \in [p]$, perform the following:

(1) Run Post-ℓ_1-quantile regression of X_a on Z^a; keep fitted value $Z^a_{-j}\widetilde{\beta}_{u,-j}$,

$$\begin{aligned}
&\widehat{\beta}_u \in \arg\min_\beta \mathbb{E}_n[K_\varpi(W)\rho_\tau(X_a - Z^a\beta)] + \lambda_u\|\beta\|_{1,\varpi}\\
&\widetilde{\beta}_u \in \arg\min_\beta \mathbb{E}_n[K_\varpi(W)\rho_\tau(X_a - Z^a\beta)] \ : \ \beta_j = 0 \ \text{ if } |\widehat{\beta}_{u,j}|\mathbb{E}_n[K_\varpi(W)(Z^a_j)^2]^{1/2} \leqslant \lambda_u.
\end{aligned}$$

(2) Run Post-Lasso of $f_u Z^a_j$ on $f_u Z^a_{-j}$; keep the residual $\widetilde{v} := f_u(Z^a_j - Z^a_{-j}\widetilde{\gamma}^j_u)$,

$$\begin{aligned}
&\widehat{\gamma}^j_u \in \arg\min_\theta \mathbb{E}_n[K_\varpi(W)f_u^2(Z^a_j - Z^a_{-j}\gamma)^2] + \lambda\|\widehat{\Gamma}_u\gamma\|_1\\
&\widetilde{\gamma}^j_u \in \arg\min_\gamma \mathbb{E}_n[K_\varpi(W)f_u^2(Z^a_j - Z^a_{-j}\gamma)^2] \ : \ \text{support}(\gamma) \subseteq \text{support}(\widehat{\gamma}^j_u).
\end{aligned}$$

(3) Run Instrumental Quantile Regression of $X_a - Z^a_{-j}\widetilde{\beta}_{u,-j}$ on Z^a_j using $\widetilde{v}$ as the instrument,

$$\check{\beta}_{u,j} \in \arg\min_{\alpha\in\mathcal{A}_{uj}} \frac{\{\mathbb{E}_n[K_\varpi(W)(1\{X_a \leqslant Z^a_j\alpha + Z^a_{-j}\widetilde{\beta}_{u,-j}\} - \tau)\widetilde{v}]\}^2}{\mathbb{E}_n[K_\varpi(W)(1\{X_a \leqslant Z^a_j\alpha + Z^a_{-j}\widetilde{\beta}_{u,-j}\} - \tau)^2\widetilde{v}^2]}$$

where $\mathcal{A}_{uj} := \{\alpha \in \mathbb{R} : |\alpha - \widetilde{\beta}_{uj}| \leqslant 10\{\mathbb{E}_n[K_\varpi(W)(Z^a_j)^2]\}^{-1/2}/\log n\}$.

A particular case is $\mathbb{C}(X_a) = \{X_a = VaR^a_\tau\}$ for a low quantile index τ, which is interpreted as with probability τ institution b is in trouble given that institution a is in trouble. They also define institution a's contribution to b as

$$\Delta CoVaR^{b|a}_\tau = CoVaR^{b|X_a=VaR^a_\tau}_\tau - CoVaR^{b|X_a=Median_a}_\tau.$$

They mainly use quantile regression to estimate the $CoVaR$ measure. More precisely, the predicted value from the quantile regression of X_b on X_a gives the value at risk of institution b conditional on institution a since VaR^b_τ given X_a is just the conditional quantile, i.e. conditional VaR

Detailed version of Algorithm 3.4 ($\mathcal{W}$-Conditional Predictive Quantile Graph Model)

For each $u = (a, \tau, \varpi) \in \mathcal{U} = V \times \mathcal{T} \times \mathcal{W}$ perform the following:

(1) Run ℓ_1-quantile regression of X_a on X_{-a} with penalty $\lambda_{0\mathcal{W}}$

$\widehat{\beta}_u \in \arg\min_\beta \mathbb{E}_n[K_\varpi(W)\rho_\tau(X_a - X'_{-a}\beta)] + \lambda_{0\mathcal{W}}\|\beta\|_{1,\varpi}$
$\widetilde{\beta}_u \in \arg\min_\beta \mathbb{E}_n[K_\varpi(W)\rho_\tau(X_a - X'_{-a}\beta)] : \ \beta_j = 0 \ \text{ if } |\widehat{\beta}_{u,j}|\mathbb{E}_n[K_\varpi(W)X_j^2]^{1/2} \leqslant \lambda_{0\mathcal{W}}$

(2) Set $\widehat{\varepsilon}_{iu} = 1\{X_{ia} \leqslant X'_{i,-a}\widetilde{\beta}_u\} - \tau$ for $i \in [n]$, $a \in V$ and $\tau \in \mathcal{T}$, $\varpi \in \mathcal{W}$. Compute the penalty level $\lambda_{\mathcal{U}}$ as the $(1-\xi)$-quantile conditional on the data of the random variable

$$\Lambda := 1.1 \max_{a\in V} \sup_{u\in\mathcal{U}} \max_{j\in[d]} \frac{|\mathbb{E}_n[gK_\varpi(W)\widehat{\varepsilon}_u X_j]|}{\sqrt{\mathbb{E}_n[K_\varpi(W)\widehat{\varepsilon}_u^2 X_j^2]}}$$

where $\{g_i : i = 1, \ldots, n\}$ is a sequence of i.i.d. standard Gaussian random variables.

(3) Run ℓ_1-quantile regression of X_a on X_{-a} with penalty $\lambda_{\mathcal{U}}$ and modified penalty loadings

$$\check{\beta}_u \in \arg\min_\beta \mathbb{E}_n[K_\varpi(W)\rho_\tau(X_a - X'_{-a}\beta)] + \lambda_{\mathcal{U}}\|\beta\|_{1,u}$$

where $\|\beta\|_{1,u} := \sum_{j\in[d]} |\beta_j|\{\mathbb{E}_n[K_\varpi(W)\widehat{\varepsilon}_u^2 X_j^2]\}^{1/2}$.

$$VaR_\tau^b|X_a = \alpha^b(\tau) + \beta^b(\tau)X_a,$$

Replacing variable X_a by its unconditional quantile, i.e. VaR_τ^a, yields

$$CoVaR_\tau^{b|X_a} = \alpha^b(\tau) + \beta^b(\tau)VaR_\tau^a \ \text{ and } \ \Delta CoVaR_\tau^{b|a} = \beta^b(\tau)(VaR_\tau^a - VaR_{50\%}^a)$$

We incorporate network spillover effects into risk measuring. We show that with QGM, individual institution's contribution to systemic risk can incorporate tail risk interconnections between institutions in the whole financial system (in the network, each node represents a financial institution now). The identified risk connections between all financial institutions constitute a systemic risk network. Note, institution a's overall systemic risk contribution, $\Delta CoVaR^{sys|a}$ measures the contribution of institution a to overall systemic risk $\sum_a \Delta CoVaR^{sys|a}$.

We define

$$Pr(X_b \leqslant CoVaR_\tau^{b|a,V\setminus\{a,b\}}|\mathbb{C}(X_a, X_{V\setminus\{a,b\}})) = \tau$$

then

$$CoVaR_\tau^{b|X_a=VaR_\tau^a, X_{V\setminus\{a,b\}}=VaR_q^{V\setminus\{a,b\}}} = \beta_0^b(\tau) + \beta_a^b(\tau)VaR_\tau^a + \beta_{V\setminus\{a,b\}}^b(\tau)VaR_\tau^{V\setminus\{a,b\}}$$

$$\Delta CoVaR_\tau^{b|a,V\setminus\{a,b\}} = \beta_a^b(\tau)(VaR_\tau^a - VaR_{50\%}^a)$$

where $\beta^b(\tau) = \{\beta_0^b(\tau), \beta_{V\setminus\{b\}}^b(\tau)\}$ is estimated via ℓ_1-penalized quantile regression.

We stack $\Delta CoVaR_{\tau}^{b|a,V\setminus\{a,b\}}$ as the (a,b)-th element of an $d\times d$ matrix $E^{\beta}(\tau)$ representing a weighted directed network of institutions. Here d is the number of total financial institutions considered. Following [5], the systemic risk contribution of firm a, $\Delta CoVaR^{sys|a}$, is the network to-degree of institution a which is defined as $\delta_a^{to} = \Delta CoVaR^{sys|a} = \sum_k \Delta CoVaR^{k|a,V\setminus\{a,k\}}$. To-degrees measure contributions of individual institutions to the overall risk of systemic network events.

Similarly, from-degree of node a is defined as $\delta_a^{from} = \Delta CoVaR^{a|sys} = \sum_b \Delta CoVaR_{\tau}^{a|b,V\setminus\{a,b\}}$. From-degrees measure exposure of individual institutions to systemic shocks from the network. The total degree δ, i.e. $\sum_a \Delta CoVaR^{sys|a}$, aggregates institution-specific systemic risk across institutions hence provides a measure of total systemic risk in the whole financial system.

Finally , we define the net contribution as $net\text{-}\Delta CoVaR^{a} = \delta_a^{to} - \delta_a^{from}$. For more about network theory, see [44].

Appendix C. Example of Simple Specifications

Next we discuss the specification and propose an estimator for CIQGMs. Although in general it is potentially hard to correctly specify coherent models, the following are examples.

Example 3 (Gaussian Case)**.** Consider the Gaussian case, $X_V \sim N(\mu,\Sigma)$ and $V=[d]$. It follows that for $a\in V$, the conditional distribution $X_a \mid X_{V\setminus\{a\}}$ satisfies

$$X_a \mid X_{V\setminus\{a\}} \sim N\left(\mu_a - \sum_{j\in V\setminus\{a\}} \frac{(\Sigma^{-1})_{aj}}{(\Sigma^{-1})_{aa}}(X_j-\mu_j), \frac{1}{(\Sigma^{-1})_{aa}}\right).$$

Therefore the conditional quantile function of X_a is linear in $X_{V\setminus\{a\}}$ and is given by

$$Q_{X_a}(\tau|X_{V\setminus\{a\}}) = \frac{\Phi^{-1}(\tau)}{(\Sigma^{-1})_{aa}^{1/2}} + \mu_a - \sum_{j\in V\setminus\{a\}} \frac{(\Sigma^{-1})_{aj}}{(\Sigma^{-1})_{aa}}(X_j-\mu_j).$$

Example 4 (Mixture of Gaussians)**.** Similar to the prior example, consider the case $X_V \mid \varpi \sim N(\mu_\varpi, \Sigma_\varpi)$ for each $\varpi\in\mathcal{W}$. It follows that for $a\in V$, the conditional distribution satisfies

$$X_a \mid X_{V\setminus\{a\}}, \varpi \sim N\left(\mu_{\varpi a} - \sum_{j\in V\setminus\{a\}} \frac{(\Sigma^{-1})_{\varpi aj}}{(\Sigma^{-1})_{\varpi aa}}(X_j-\mu_{\varpi j}), \frac{1}{(\Sigma^{-1})_{\varpi aa}}\right).$$

Again the conditional quantile function of X_a is linear in $X_{V\setminus\{a\}}$ and is given by

$$Q_{X_a}(\tau|X_{V\setminus\{a\}},\varpi) = \frac{\Phi^{-1}(\tau)}{(\Sigma^{-1})_{\varpi aa}^{1/2}} + \mu_{\varpi a} - \sum_{j\in V\setminus\{a\}} \frac{(\Sigma^{-1})_{\varpi aj}}{(\Sigma^{-1})_{\varpi aa}}(X_j-\mu_{\varpi j}).$$

Example 5 (Monotone Transformations)**.** Consider the Gaussian case, $X_j = h_j(Y_j)$ where $Y_V \sim N(\mu,\Sigma)$, $j\in V=[d]$. It follows that for $a\in V$, the conditional quantile function satisfies

$$Q_{X_a}(\tau|X_{V\setminus\{a\}}) = h_a\left(\frac{\Phi^{-1}(\tau)}{(\Sigma^{-1})_{aa}^{1/2}} + \mu_a - \sum_{j\in V\setminus\{a\}} \frac{(\Sigma^{-1})_{aj}}{(\Sigma^{-1})_{aa}}(h_j^{-1}(X_j)-\mu_j)\right).$$

In particular if $(h_j : j\in V)$ are monotone polynomials, the expression above is a sum of monomials with fractional and integer exponents.

Example 6 (Multiplicative Error Model)**.** Consider $d = 2$ so that $V = \{1, 2\}$. Assume that X_2 and ε are independent positive random variables. Assume further that they relate to X_1 as

$$X_1 = \alpha + \varepsilon X_2.$$

In this case we have that the conditional quantile functions are linear and given by

$$Q_{X_1}(\tau|X_2) = \alpha + F_\varepsilon^{-1}(\tau)X_2 \quad \text{and} \quad Q_{X_2}(\tau|X_1) = (X_1 - \alpha)/F_\varepsilon^{-1}(1 - \tau).$$

■

Example 7 (Additive Error Model)**.** Consider $d = 2$ so that $V = \{1, 2\}$. Let $X_2 \sim U(0, 1)$ and $\epsilon \sim U(0, 1)$ be independent random variables. Also define the random variable X_1 is defined as

$$X_1 = \alpha + \beta X_2 + \varepsilon.$$

It follows that $Q_{X_1}(\tau|X_2) = \alpha + \beta X_2 + \tau$. However, if $\beta = 0$, we have $Q_{X_2}(\tau|X_1) = \tau$, and for $\beta > 0$, direct calculations yield that

$$Q_{X_2}(\tau|X_1) = \begin{cases} \frac{\tau}{\beta}(X_1 - \alpha), & \text{if} \quad X_1 \leqslant \alpha + \beta \\ \tau + (1 - \tau)(X_1 - \alpha - \beta), & \text{if} \quad X_1 \geqslant \alpha + \beta \end{cases}$$

where we note that $X_1 \in [\alpha, 1 + \alpha + \beta]$. ■

Although a linear specification is correct for Examples 3 and 6, Example 7 above illustrates that we need to consider more general transformation of the basic covariates X_V in the specification for each conditional quantile function. Nonetheless, specifications with additional non-linear terms can approximate non-drastic departures from normality.

Appendix D. Proofs of Section 4

Proof of Theorem 1. By Lemma 6, under Condition CI, for any θ such that $\|\theta\|_0 \leqslant Cs\ell_n$, $\ell_n \to \infty$ slowly, we have that

$$\|\sqrt{f_u}Z^a\theta\|_{n,\varpi}/\{\mathrm{E}[K_\varpi(W)f_u(Z^a\theta)^2]\}^{1/2} = 1 + o_P(1).$$

Moreover, $\mathrm{E}[K_\varpi(W)f_u(Z^a\theta)^2] \geqslant \underline{f}_u\mathrm{E}[K_\varpi(W)(Z^a\theta)^2]$, $\mathrm{E}[K_\varpi(W)(Z^a\theta)^2] = \mathrm{E}[(Z^a\theta)^2 \mid \varpi]\mathrm{P}(\varpi)$, and $\mathrm{E}[(Z^a\theta)^2 \mid \varpi] \geqslant c\|\theta\|^2$ by Condition CI. Lemma 6 further imply that the ratio of the minimal and maximal eigenvalues of order $s\ell_n$ are bounded away from zero and from above uniformly over $\varpi \in \mathcal{W}$ and $a \in V$ with probability $1 - o(1)$. Therefore, since $c\{\mathrm{P}(\varpi)\}^{1/2}\|\delta\|_1 \leqslant \|\delta\|_{1,\varpi} \leqslant C\{\mathrm{P}(\varpi)\}^{1/2}\|\delta\|_1$, we have $\kappa_{u,2\mathbf{c}} \geqslant c$ uniformly over $u \in \mathcal{U}$ with the same probability.

Consider the events Ω_1, Ω_2, and Ω_3 as defined in (E.33), (E.34) and (E.35). By the choice of λ_u we have $\mathrm{P}(\Omega_1) \geqslant 1 - o(1)$, and $\mathrm{P}(\Omega_2) \geqslant 1 - o(1)$ by Condition CI with $R_{u\gamma} \leqslant Cs\log(p|V|n)/n$ by Lemma 2, and $\mathrm{P}(\Omega_3) \geqslant 1 - \xi$ by Lemma 3 with $t_3 \leqslant Cn^{-1/2}\sqrt{(1 + d_W)\log(p|V|nL_f/\xi)}$. Finally, we have $q_{A_u} \geqslant c\{\sqrt{s}\max_{i\leqslant n}\|Z_i^a\|_\infty\}^{-1} \geqslant c'\{\sqrt{s}M_n\}^{-1}$ since we can assume $\|\sqrt{f_u}Z^a\delta_u\|_{n,\varpi} \geqslant c\sqrt{s\log(p|V|n)/n}$. Therefore, setting $\xi = 1/\log n$, by Lemma 1 we have uniformly over $u \in \mathcal{U}$

$$\begin{aligned} \|\sqrt{f_u}Z^a(\widehat{\beta}_u - \beta_u)\|_{n,\varpi} &\lesssim \sqrt{(1 + (t_3/\lambda_u)R_{u\gamma}} + (\lambda_u + t_3)\sqrt{s} \lesssim \sqrt{\tfrac{s(1+d_W)\log(p|V|n)}{n\tau(1-\tau)}} \\ \|\widehat{\beta}_u - \beta_u\|_{1,\varpi} &\lesssim s\sqrt{\tfrac{(1+d_W)\log(p|V|n)}{n}} \end{aligned} \tag{D.28}$$

where we used that $\lambda_u \leqslant C\sqrt{\frac{(1+d_W)\log(p|V|n)}{n}}$ by Lemma 15 to bound $\Lambda_a(1-\gamma/\{|V|n^{1+2d_W}\} \mid X_{-a}, W)$ under $M_n^2\log(p|V|n/\{\tau(1-\tau)\}) = o(n\tau(1-\tau)\mu_{\mathcal{W}})$ for all $\tau \in \mathcal{T}$.

Let $\delta_u = \widehat{\beta}_u - \beta_u$. By triangle inequality it follows that

$$\{\mathrm{E}[K_\varpi(W)f_u(Z^a\delta_u)^2]\}^{1/2} \leqslant \|\sqrt{f_u}Z^a\delta_u\|_{n,\varpi} + \|\delta_u\|_{1,\varpi}\{|(\mathbb{E}_n - \mathrm{E})[K_\varpi(W)f_u(Z^a\delta_u)^2/\|\delta_u\|_{1,\varpi}^2]\}^{1/2} \tag{D.29}$$

and the last term can be bounded by

$$\begin{aligned} \sup_{\|\delta\|_{1,\varpi}\leqslant 1} |(\mathbb{E}_n - \mathrm{E})[K_\varpi(W)f_u(Z^a\delta)^2/\|\delta\|_{1,\varpi}^2]| &\leqslant |(\mathbb{E}_n - \mathrm{E})[K_\varpi(W)f_u\{Z_k^a/\widehat{\sigma}_{\varpi k}\}\{Z_j^a/\widehat{\sigma}_{\varpi j}\}]| \\ &\lesssim \tfrac{1}{\mathrm{P}(\varpi)}\sqrt{\tfrac{(1+d_W)\log(p|V|n)}{n}}. \end{aligned}$$

Combining the relations above with (D.29), under $(1+d_W)s^2\log(p|V|n) = o(n)$ we have uniformly over $u \in \mathcal{U}$

$$\begin{aligned} \|\delta_u\| &\lesssim \{\mathrm{E}[(Z^a\delta_u)^2 \mid \varpi]\}^{1/2} \lesssim \{\mathrm{P}(\varpi)\}^{-1/2}\{\mathrm{E}[K_\varpi(W)(Z^a\delta_u)^2]\}^{1/2} \\ &\lesssim \{\mathrm{P}(\varpi)\underline{f}_u\}^{-1/2}\{\mathrm{E}[K_\varpi(W)f_u(Z^a\delta_u)^2]\}^{1/2} \\ &\leqslant \{\mathrm{P}(\varpi)\underline{f}_u\}^{-1/2}\|\sqrt{f_u}Z^a\delta_u\|_{n,\varpi} + \{\mathrm{P}(\varpi)\underline{f}_u\}^{-1/2}\sqrt[4]{\tfrac{(1+d_W)\log(p|V|n)}{n}}\|\delta_u\|_{1,\varpi} \\ &\leqslant C\sqrt{\tfrac{s(1+d_W)\log(p|V|n)}{n\underline{f}_u\mathrm{P}(\varpi)}}. \end{aligned}$$

Finally, let $\widehat{\beta}_u^{\bar{\lambda}}$ be obtained by thresholding the estimator $\widehat{\beta}_u$ with $\bar{\lambda} := \sqrt{(1+d_W)\log(p|V|n)/n}$ (note that each component is weighted by $\mathbb{E}_n[K_\varpi(W)(Z_j^a)^2]^{1/2}$). By Lemma 16, we have

$$\begin{aligned} \|Z^a(\widehat{\beta}_u^{\bar{\lambda}} - \beta_u)\|_{n,\varpi} &\lesssim \sqrt{s(1+d_W)\log(p|V|n)/n} \\ \|\widehat{\beta}_u^{\bar{\lambda}} - \beta_u\|_{1,\varpi} &\lesssim s\sqrt{(1+d_W)\log(p|V|n)/n} \\ |\mathrm{support}(\widehat{\beta}_u^{\bar{\lambda}})| &\lesssim s \end{aligned}$$

by the choice of $\bar{\lambda}$ and the rates in (D.28)

∎

Proof of Theorem 2. We verify Assumption 4 and Condition WL for the weighted Lasso model with index set $\mathcal{U}\times[p]$ where $Y_u = K_\varpi(W)Z_j^a$, $X_u = K_\varpi(W)Z_{-j}^a$, $\theta_u = \bar{\gamma}_u^j$, $a_u = (f_u, \bar{r}_{uj})$, $\bar{r}_{uj} = K_\varpi(W)Z_{-j}^a(\gamma_u^j - \bar{\gamma}_u^j)$, $S_{uj} = K_\varpi(W)f_u^2(Z_j^a - Z_{-j}^a\gamma_u^j)Z_{-j}^a = K_\varpi(W)f_u v_{uj}Z_{-j}^a$, and $w_u = K_\varpi(W)f_u^2$. We will take $N_n = |V|p^2\{pn^3\}^{1+d_W}$ in the definition of λ.

We first verify Condition WL. We have $\mathrm{E}[S_{ujk}^2] \leqslant \bar{f}^2\mathrm{E}[|v_{uj}Z_{-jk}^a|^2] \leqslant \mathrm{E}[|v_{uj}|^4 + |Z_{-jk}^a|^4] \leqslant C$ by the bounded fourth moment condition. We have that

$$\frac{\mathrm{E}[|S_{ujk}|^3]^{1/3}}{\mathrm{E}[|S_{ujk}|^2]^{1/2}} = \frac{\mathrm{E}[|S_{ujk}|^3 \mid \varpi]^{1/3}}{\mathrm{E}[|S_{ujk}|^2 \mid \varpi]^{1/2}}\{\mathrm{P}(\varpi)\}^{-1/6} = \frac{\mathrm{E}[|f_u v_{uj}Z_{-jk}^a|^3 \mid \varpi]^{1/3}}{\mathrm{E}[|f_u v_{uj}Z_{-jk}^a|^2 \mid \varpi]^{1/2}}\{\mathrm{P}(\varpi)\}^{-1/6} =: M_{uk}$$

By the choice N_n and since $\Phi^{-1}(1-t) \leqslant C\sqrt{\log(1/t)}$, we have $M_{uk}\Phi^{-1}(1-\gamma/2pN_n) \leqslant M_{uk}C(1+d_W)\log^{1/2}(pn|V|) \leqslant C\delta_n n^{1/6}$ where the last inequality holds by Condition CI so Condition WL(i) holds.

To verify Condition WL(ii) we will establish the validity of the choice of N_n and use that

$$S_{ujk} - S_{u'jk} = \{K_\varpi(W)f_u^2 - K_{\varpi'}(W)f_{u'}^2\}\{Z_j^a - Z_{-j}^a\gamma_u^j\}Z_k^a + K_{\varpi'}(W)f_{u'}^2\{Z_{-j}^a(\gamma_u^j - \gamma_{u'}^j)\}Z_k^a.$$

We have that $|K_\varpi(W)f_u^2 - K_{\varpi'}(W)f_{u'}^2| \leqslant K_\varpi(W)K_{\varpi'}(W)|f_u^2 - f_{u'}^2| + (f_u + f_{u'})^2|K_\varpi(W) - K_{\varpi'}(W)|$. We will consider $u = (a, \tau, \varpi)$ and $u' = (a, \tau', \varpi')$. By Condition CI we have that for some $\bar{L}$ such that $\bar{L} \lesssim \log(p|V|n)$

$$|f_u - f_{u'}| \leqslant \bar{L}\|u - u'\| \ \text{ and } \ \mathrm{E}[|K_\varpi(W) - K_{\varpi'}(W)|] \leqslant \bar{L}\|\varpi - \varpi'\|. \tag{D.30}$$

Further, by Lemma 5 we have

$$\|\gamma_u^j - \gamma_{u'}^j\| \leqslant L_\gamma\{\|u - u'\| + \|\varpi - \varpi'\|^{1/2}\} \tag{D.31}$$

for some L_γ satisfying $\log(L_\gamma) \leqslant C\log(|V|pn)$. It follows that since $f_u + f_{u'} \leqslant 2f_u + L\|u - u'\|$

$$\begin{array}{rl} |\mathbb{E}_n[S_{ujk} - S_{u'jk}]| & \leqslant \mathbb{E}_n[|S_{ujk} - S_{u'jk}|] \\ & \leqslant \mathbb{E}_n[K_\varpi(W)K_{\varpi'}(W)|f_u - f_{u'}|] \max_{i\leqslant n}(2f_{ui} + \bar{L}\|u - u'\|)|Z_{ij}^a - Z_{i,-j}^a\gamma_u^j| \ \|Z_i^a\|_\infty \\ & + \mathbb{E}_n[|K_\varpi(W) - K_{\varpi'}(W)|] \max_{i\leqslant n}(2f_{ui} + L\|u - u'\|)|Z_{ij}^a - Z_{i,-j}^a\gamma_u^j| \ \|Z_i^a\|_\infty \\ & + \bar{f}^2\sqrt{p}\max_{i\leqslant n}\|Z_i^a\|_\infty^2\|\gamma_u^j - \gamma_{u'}^j\| \\ & \leqslant \bar{L}\|u - u'\|\max_{i\leqslant n}|v_{uji}| \ \|Z^a\|_\infty + \bar{L}^2\|u - u'\|^2\sqrt{p}\|\gamma_u^j\|\max_{i\leqslant n}\|Z_i^a\|_\infty^2 \\ & + \mathbb{E}_n[|K_\varpi(W) - K_{\varpi'}(W)|]\max_{i\leqslant n}|2v_{uji}| \ \|Z_i^a\|_\infty + \bar{L}\|u - u'\|\sqrt{p}\|\gamma_u^j\|\max_{i\leqslant n}\|Z_i^a\|_\infty^2 \\ & + L_\gamma\{\|u - u'\| + \|\varpi - \varpi'\|^{1/2}\}\bar{f}^2\sqrt{p}\max_{i\leqslant n}\|Z_i^a\|_\infty^2 \end{array}$$

.

Note that $\max_{i\leqslant n}|v_{uji}| \ \|Z^a\|_\infty + \max_{i\leqslant n}\|Z^a\|_\infty^2 \lesssim_P n^{2/q}M_n$. For $d_{\mathcal{U}} = \|\cdot\|$, an uniform ϵ-cover of $\mathcal{U}$ satisfies $(6\mathrm{diam}(\mathcal{U})/\epsilon)^{1+d_W} \geqslant N(\epsilon, \mathcal{U}, \|\cdot\|)$. Taking $1/\epsilon = \{\bar{L} + L_\gamma^2\}pn^2M_n^2\log^2(p|V|n)/\{\mu_{\mathcal{W}}\underline{f}_{\mathcal{U}}^2\} \leqslant pn^3$ we have with probability $1 - o(1)$ that

$$\begin{array}{rl} |\mathbb{E}_n[S_{ujk} - S_{u'jk}]| & \lesssim \bar{L}^2\epsilon^2\sqrt{p}n^{2/q}M_n\log n + \bar{L}\epsilon\sqrt{p}n^{2/q}M_n^2\log n + \epsilon^{1/2}L_\gamma\sqrt{p}n^{2/q}M_n^2\log n \\ & + \mathbb{E}_n[|K_\varpi(W) - K_{\varpi'}(W)|]n^{2/q}M_n\log n \\ & \lesssim \delta_n n^{-1/2}\{\mu_{\mathcal{W}}\underline{f}_{\mathcal{U}}^2\}^{1/2} + \mathbb{E}_n[|K_\varpi(W) - K_{\varpi'}(W)|]n^{2/q}M_n\log n \end{array}$$

since $\sup_{u\in\mathcal{U}, j\in[p]}\|\gamma_u^j\| \leqslant C$ and $\bar{f} \leqslant C$. Next note that with probability $1 - o(1)$

$$\begin{array}{rl} \mathbb{E}_n[|K_\varpi(W) - K_{\varpi'}(W)|] & \leqslant |(\mathbb{E}_n - \mathrm{E})[|K_\varpi(W) - K_{\varpi'}(W)|]| + \mathrm{E}[|K_\varpi(W) - K_{\varpi'}(W)|] \\ & \leqslant \displaystyle\sup_{\varpi,\varpi'\in\mathcal{W}, \|\varpi - \varpi'\|\leqslant\epsilon} |(\mathbb{E}_n - \mathrm{E})[|K_\varpi(W) - K_{\varpi'}(W)|]| + \bar{L}\epsilon \\ & \lesssim \sqrt{\frac{d_W\log(n/\epsilon)}{n}}\epsilon^{1/2} + \frac{d_W\log(n/\epsilon)}{n} + \bar{L}\epsilon \end{array}$$

which yields uniformly over $u \in \mathcal{U}$ and $j \in [p]$

$$|\mathbb{E}_n[S_{ujk} - S_{u'jk}]| \lesssim \delta_n n^{-1/2}\mu_{\mathcal{W}}^{1/2}\underline{f}_{\mathcal{U}}$$

under $\sqrt{\epsilon d_W\log(n/\epsilon)}n^{2/q}M_n\log n = o(\mu_{\mathcal{W}}^{1/2}\underline{f}_{\mathcal{U}})$ and $d_W\log(n/\epsilon)n^{2/q}M_n\log n = o(n^{1/2}\mu_{\mathcal{W}}^{1/2}\underline{f}_{\mathcal{U}})$. This implies

$$\sup_{|u-u'|\leqslant\epsilon}\max_{j\in[p],k\in[p-1]}\frac{|\mathbb{E}_n[S_{ujk} - S_{u'jk}]|}{\mathrm{E}[S_{ujk}^2]^{1/2}} \leqslant \delta_n n^{-1/2}$$

since $\mathrm{E}[S_{ujk}^2] \geqslant c\mu_{\mathcal{W}}\underline{f}_{\mathcal{U}}^2$. The same choice of ϵ also implies

$$\sup_{|u-u'|\leqslant\epsilon}\max_{j,k\in[p]}\frac{|\mathrm{E}[S_{ujk}^2 - S_{u'jk}^2]|}{\mathrm{E}[S_{ujk}^2]} \leqslant \delta_n$$

To establish the last requirement of Condition WL(ii), we will apply Lemma 2 with the vector $\{(\bar{X})_{uj} = S_{uj}, u \in \mathcal{U}^\epsilon, j \in [p]\}$, $\widehat{\mathcal{U}} := \mathcal{U}^\epsilon \times [p]$, and since

$$K^2 = \mathrm{E}[\max_{i \leqslant n} \sup_{u \in \mathcal{U}} \max_{j \in [p], k \in [p-1]} S^2_{ujk}] \leqslant \mathrm{E}[\max_{i \leqslant n} \sup_{u \in \mathcal{U}, j \in [p]} |v_{i,uj}|^2 \| f_{ui} Z^a_i \|^2_\infty] \leqslant \bar{f}^2 n^{4/q} M_n^2 L_n^2$$

we have

$$\begin{aligned} \sup_{u \in \mathcal{U}} \max_{j \in [p], k \in [p-1]} |(\mathbb{E}_n - \mathrm{E})[S^2_{ujk}]| & \leqslant \sup_{u \in \mathcal{U}^\epsilon} \max_{j \in [p], k \in [p-1]} |(\mathbb{E}_n - \mathrm{E})[S^2_{ujk}]| + \Delta_n \\ & \leqslant C n^{-1/2} n^{2/q} M_n L_n \log^{1/2}(p|V|n) \leqslant C \delta_n \mu_{\mathcal{W}} \underline{f}_{\mathcal{U}} + \Delta_n \end{aligned}$$

under $n^{4/q} M_n^2 L_n^2 \log(p|V|n) \leqslant \delta_n n \mu^2_{\mathcal{W}} \underline{f}^2_{\mathcal{U}}$ where

$$\Delta_n := \sup_{u,u' \in \mathcal{U}, \|u-u'\| \leqslant \epsilon} \max_{j \in [p], k \in [p-1]} |(\mathbb{E}_n - \mathrm{E})[S^2_{ujk}] - (\mathbb{E}_n - \mathrm{E})[S^2_{u'jk}]|.$$

Note that

$$\begin{aligned} |(\mathbb{E}_n - \mathrm{E})[S^2_{ujk}] - (\mathbb{E}_n - \mathrm{E})[S^2_{u'jk}]| & \leqslant |\mathbb{E}_n[S^2_{ujk} - S^2_{u'jk}]| + |\mathrm{E}[S^2_{ujk} - S^2_{u'jk}]| \\ & \leqslant \mathbb{E}_n[|S_{ujk} - S_{u'jk}|] \sup_{u \in \mathcal{U}} \max_{i \leqslant n} |2 S_{u'jk,i}| + |\mathrm{E}[S^2_{ujk} - S^2_{u'jk}]| \\ & \lesssim \delta_n n^{-1/2} \{\mu^{1/2}_{\mathcal{W}} \underline{f}_{\mathcal{U}}\} \bar{f} \sup_{u \in \mathcal{U}} \max_{i \leqslant n} |v_{uji}| \|Z^a\|_\infty + \delta_n \mu_{\mathcal{W}} \underline{f}^2_{\mathcal{U}} \\ & \lesssim \delta_n n^{-1/2} \{\mu^{1/2}_{\mathcal{W}} \underline{f}_{\mathcal{U}}\} n^{2/q} M_n L_n \log n + \delta_n \mu_{\mathcal{W}} \underline{f}^2_{\mathcal{U}} \end{aligned}$$

with probability $1 - o(1)$ where we used the previous two results. Therefore $\Delta_n \lesssim \delta_n \mu_{\mathcal{W}} \underline{f}^2_{\mathcal{U}}$ as required.

To verify Assumption 4(a), note that $[\partial_\theta M_u(Y_u, X, \theta_u) - \partial_\theta M_u(Y_u, X, \theta_u, a_u)]'\delta = -f_u^2 K_\varpi(W) \bar{r}_{uj} Z^a_{-j} \delta$, so that by Cauchy-Schwartz, we have

$$\mathbb{E}_n[\partial_\theta M_u(Y_u, X, \theta_u) - \partial_\theta M_u(Y_u, X, \theta_u, a_u)]'\delta \quad \leqslant \|f_u \bar{r}_{uj}\|_{\varpi,2} \|f_u Z^a_{-j} \delta\|_{\varpi,2} \leqslant C_{un} \|f_u Z^a_{-j} \delta\|_{\varpi,2}$$

where we choose C_{un} so that $\{C_{un} \geqslant \max_{j \in [p]} \|f_u \bar{r}_{uj}\|_{\varpi,2} : u \in \mathcal{U}\}$ with probability $1 - o(1)$. By Lemma 4, uniformly over $u \in \mathcal{U}, j \in [\tilde{p}]$ we have

$$\|f_u \bar{r}_{uj}\|_{\varpi,2} = \|f_u Z^a_{-j} (\gamma^j_u - \bar{\gamma}^j_u)\|_{\varpi,2} \lesssim \underline{f}_u \{\mathrm{P}(\varpi)\}^{1/2} \{n^{-1} s \log(p|V|n)\}^{1/2}$$

so that $C_{un} = \underline{f}_u \{\mathrm{P}(\varpi)\}^{1/2} \{n^{-1} s \log(p|V|n)\}^{1/2}$.

Next we show that Assumption 4(b) holds. First note the uniform convergence of the loadings

$$\sup_{u \in \mathcal{U}, j \in [p], k \in [p]} |\mathbb{E}_n[S^2_{ujk}] - \mathrm{E}[S^2_{ujk}]| + |(\mathbb{E}_n - \mathrm{E})[K_\varpi(W) f_u^2 |Z^a_j Z^a_{-jk}|^2]| \leqslant \delta_n \mu_{\mathcal{W}} \underline{f}^2_{\mathcal{U}}$$

so that $\mathbb{E}_n[S^2_{ujk}]/\mathrm{E}[S^2_{ujk}] = 1 + o_P(1)$. It follows that $\tilde{\mathbf{c}}$ is bounded above by a constant for n large enough. Indeed, uniformly over $u \in \mathcal{U}$, $j \in [p]$, since $c\underline{f}_u \leqslant \mathrm{E}[|f_u v_{uj} Z^a_k|^2 \mid \varpi]^{1/2} \leqslant C \underline{f}_u$, with probability $1 - o(1)$ we have $c\underline{f}_u \mathrm{P}(\varpi)^{1/2} \leqslant \widehat{\Psi}_{u0jj} \leqslant C \underline{f}_u \mathrm{P}(\varpi)^{1/2}$ so that $c/C \leqslant \|\widehat{\Psi}_{u0}\|_\infty \|\widehat{\Psi}^{-1}_{u0}\|_\infty \leqslant C/c$.

Assumption 4(c) follows directly from the choice of $M_u(Y_u, X_u, \theta) = K_\varpi(W) f_u^2 (Z^a_j - Z^a_{-j}\theta)^2$ with $\bar{q}_{A_u} = \infty$.

The result for the rate of convergence then follows from Lemma 21, namely

$$\|f_u X'_u (\widehat{\gamma}^j_u - \gamma^j_u)\|_{n,\varpi} \lesssim \frac{\|\widehat{\Psi}_{u0}\|_\infty}{\bar{\kappa}_{u,2\mathbf{c}}} \sqrt{\frac{s \log(p|V|n)}{n}} + C_{un} \lesssim \frac{\underline{f}_u \mathrm{P}(\varpi)^{1/2}}{\bar{\kappa}_{u,2\mathbf{c}}} \sqrt{\frac{s \log(p|V|n)}{n}} \tag{D.32}$$

By Lemma 6 we have that for sparse vectors, $\|\theta\|_0 \leqslant \ell_n s$ satisfies

$$\|f_u Z^a_{-j}\theta\|^2_{n,\varpi}/\mathrm{E}[K_\varpi(W) f_u^2\{Z^a_{-j}\theta\}^2] = 1 + o_P(1)$$

so that $\phi_{\max}(\ell_n s, uj) \leqslant C\underline{f}_u^2 \mathrm{P}(\varpi)$ and $\widehat{s}_{uj} \leqslant \min_{m\in\mathcal{M}_u} \phi_{\max}(m, uj) L_u^2 \leqslant Cs$ provided $L_u^2 \lesssim s\{\underline{f}_u^2\mathrm{P}(\varpi)\}^{-1}$. Indeed, with probability $1-o(1)$, we have $\|\widehat{\Psi}_{u0}^{-1}\|_\infty \leqslant C\underline{f}_u^{-1}\{\mathrm{P}(\varpi)\}^{-1/2}$, so that $L_u \lesssim \underline{f}_u^{-1}\mathrm{P}(\varpi)^{-1/2}\frac{n}{\lambda}\{C_{un}+L_{un}\}$. Moreover, we can take $C_{un} \lesssim \underline{f}_u\{\mathrm{P}(\varpi)n^{-1}s\log(p|V|n)\}^{1/2}$, and $L_{un} \lesssim \{n^{-1}s\log(p|V|n)\}^{1/2}$ in Assumption C4 because

$$\begin{aligned}
&|\{\mathbb{E}_n[\partial_\gamma M_u(Y_u, X_u, \widehat{\gamma}^j_u) - \partial_\gamma M_u(Y_u, X_u, \gamma^j_u)]\}'\delta| \\
&= 2|\{\mathbb{E}_n[K_\varpi(W) f_u^2\{X_u'(\widehat{\gamma}^j_u - \gamma^j_u)\} X_u'\delta| \\
&\leqslant 2\|f_u X_u'(\widehat{\gamma}^j_u - \gamma^j_u)\|_{n,\varpi}\|f_u X_u'\delta\|_{n,\varpi} =: L_{un}\|f_u X_u'\delta\|_{n,\varpi},
\end{aligned}$$

where the last inequality hold by (D.32) since $\bar{\kappa}_{u,2\mathbf{c}} \geqslant c\underline{f}_u\{\mathrm{P}(\varpi)\}^{1/2}$. The latter holds since for any $\|\delta\| = 1$, we have

$$\begin{aligned}
c\underline{f}_u \mathrm{P}(\varpi) &\leqslant \mathrm{E}[K_\varpi(W) f_u (Z^a\delta)^2] \\
&\leqslant \{\mathrm{E}[K_\varpi(W) f_u^2 (Z^a\delta)^2]\}^{1/2}\{\mathrm{E}[K_\varpi(W)(Z^a\delta)^2]\}^{1/2} \\
&\leqslant \{\mathrm{E}[K_\varpi(W) f_u^2 (Z^a\delta)^2]\}^{1/2} C\{\mathrm{P}(\varpi)\}^{1/2}
\end{aligned}$$

where the first inequality follows from the definition of $\underline{f}_u$, $\|\delta\| = 1$, and Condition CI, so that we have $\{\mathrm{E}[K_\varpi(W) f_u^2 (Z^a\delta)^2]\}^{1/2} \geqslant c'\underline{f}_u\{\mathrm{P}(\varpi)\}^{1/2}$.

The sparsity result follows from Lemma 20. The result for Post Lasso follows from Lemma 19 under the growth requirements in Condition CI.

■

Proof of Theorem 3. We will verify Assumptions 1 and 2, and the result follows from Theorem 5. The estimate of the nuisance parameter is constructed from the estimators in Steps 1 and 2 of the Algorithm.

For each $u = (a, \tau, \varpi) \in \mathcal{U}$ and $j \in [p]$, let $W_{uj} = (W, X_a, Z^a, v_{uj}, r_u)$, where $v_{uj} = f_u(Z^a_j - Z^a_{-j}\gamma^j_u)$ and let $\theta_{uj} \in \Theta_{uj} = \{\theta \in \mathbb{R} : |\theta - \beta_{uj}| \leqslant c/\log n\}$ (Assumption 1(i) holds). The score function as

$$\psi_{uj}(W_{uj}, \theta, \eta_{uj}) = K_\varpi(W)\{\tau - 1\{X_a \leqslant Z^a_j\theta + Z^a_{-j}\beta_{u,-j} + r_u\}\} f_u^2 (Z^a_j - Z^a_{-j}\gamma^j_u)$$

where the nuisance parameter is $\eta_{uj} = (\beta^j_u, \gamma^j_u, r_u)$ where the last component is a function $r_u = r_u(X)$. Recall that $K_\varpi(W) \in \{0, 1\}$. For $a_n = \max(n, p, |V|)$ let $\mathcal{H}_{uj} = \{\eta = (\eta^{(1)}, \eta^{(2)}, \eta^{(3)}) : \|\eta - \eta_{uj}\|_e \leqslant \tau_n$ where $\|\eta - \eta_{uj}\|_e = \|(\delta^{(1)}_\eta, \delta^{(2)}_\eta, \delta^{(3)}_\eta)\|_e = \max\{\|\delta^{(1)}_\eta\|, \|\delta^{(2)}_\eta\|, \mathrm{E}[|\delta^{(3)}_\eta|^2]^{1/2}\}$, and

$$\tau_n := C \sup_{u\in\mathcal{U}} \frac{1}{1 \wedge \underline{f}_{\mathcal{U}}} \sqrt{\frac{s\log a_n}{n\mu_{\mathcal{W}}}}$$

The differentiability of the mapping $(\theta, \eta) \in \Theta_{uj} \times \mathcal{H}_{uj} \mapsto \mathrm{E}\psi_{uj}(W_{uj}, \theta, \eta)$ follows from the differentiability of the conditional probability distribution of X_a given $X_{V\setminus a}$ and ϖ. Let $\eta = (\eta^{(1)}, \eta^{(2)}, \eta^{(3)})$, $\delta_\eta = (\delta^{(1)}_\eta, \delta^{(2)}_\eta, \delta^{(3)}_\eta)$, and $\theta_{\bar{r}} = \theta + \bar{r}\delta_\theta$, $\eta_{\bar{r}} = \eta + \bar{r}\delta_\eta$.

To verify Assumption 1(v)(a) with $\alpha = 1$, for any $(\theta,\eta),(\bar\theta,\bar\eta) \in \Theta_{uj}\times\mathcal{H}_{uj}$ note that since $f_{X_a|X_{-a},\varpi}$ is uniformly bounded from above by $\bar f$, therefore

$$\begin{aligned}
&\mathrm{E}[\{\psi_{uj}(W_{uj},\theta,\eta)-\psi_{uj}(W_{uj},\bar\theta,\bar\eta)\}^2]^{1/2}\\
&\leqslant \mathrm{E}[|Z^a_{-j}(\eta^{(2)}-\bar\eta^{(2)})|^2]^{1/2}+\mathrm{E}[(Z^a_j-Z^a_{-j}\bar\eta^{(2)})^2\bar f\{|\eta^{(3)}-\bar\eta^{(3)}|+|Z^a_{-j}(\eta^{(1)}-\bar\eta^{(1)})|+|Z^a_j(\theta-\bar\theta)|\}]^{1/2}\\
&\leqslant C\|\eta^{(2)}-\bar\eta^{(2)}\|+\bar f^{1/2}\mathrm{E}[(Z^a_j-Z^a_{-j}\bar\eta^{(2)})^4]^{1/4}\{\mathrm{E}[|\eta^{(3)}-\bar\eta^{(3)}|^2]^{1/4}+C\|\eta^{(1)}-\bar\eta^{(1)}\|+|\theta-\bar\theta|\}^{1/2}\\
&\leqslant C'|\theta-\bar\theta|^{1/2}\vee\|\eta-\bar\eta\|_e^{1/2}
\end{aligned}$$

for some constance $C'<\infty$ since by Condition CI we have $\mathrm{E}[|Z^a\xi|^4]^{1/4}\leqslant C\|\xi\|$ for all vectors ξ, and the conditions $\sup_{u\in\mathcal{U},j\in[p]}\|\gamma^j_u\|\leqslant C$, $\sup_{\theta\in\Theta_{uj}}|\theta|\leqslant C$, and $\sqrt{s\log(a_n)}\leqslant\delta_n\sqrt{n}$. This implies that $\|\eta^{(2)}-\bar\eta^{(2)}\|\leqslant\|\eta^{(2)}-\eta^{(2)}_{uj}\|+\|\eta^{(2)}_{uj}-\bar\eta^{(2)}\|\leqslant 1$ so that $\|\eta^{(2)}-\bar\eta^{(2)}\|\leqslant\|\eta^{(2)}-\bar\eta^{(2)}\|^{1/2}$.

To verify Assumption 1(v)(b), let $t_{\bar r}=Z^a_j\theta_{\bar r}+Z^a_{-j}\eta^{(1)}_{\bar r}+\eta^{(3)}_{\bar r}$. We have

$$\begin{aligned}
&\partial_r\mathrm{E}(\psi_{uj}(W_{uj},\theta+r\delta_\theta,\eta+r\delta_\eta))\big|_{r=\bar r}=\\
&-\mathrm{E}[K_\varpi(W)f_{X_a|X_{-a},\varpi}(t_{\bar r})(Z^a_j-Z^a_{-j}\eta^{(2)}_{\bar r})\{Z^a_j\delta_\theta+Z^a_{-j}\delta^{(1)}_\eta+\delta^{(3)}_\eta\}]\\
&-\mathrm{E}[K_\varpi(W)\{u-F_{X_a|X_{-a},\varpi}(t_{\bar r})\}Z^a_{-j}\delta^{(2)}_\eta]
\end{aligned}$$

Applying Cauchy-Schwartz we have that

$$\begin{aligned}
&\big|\partial_r\mathrm{E}(\psi_{uj}(W_{uj},\theta+r\delta_\theta,\eta+r\delta_\eta))|_{r=\bar r}\big|\\
&\leqslant \bar f\mathrm{E}[(Z^a_j-Z^a_{-j}\eta^{(2)}_{\bar r})^2]^{1/2}\{\mathrm{E}[(Z^a_j)^2]^{1/2}|\delta_\theta|+\mathrm{E}[(Z^a_{-j}\delta^{(1)}_\eta)^2]^{1/2}+\mathrm{E}[|\delta^{(3)}_\eta|^2]^{1/2}\}+\mathrm{E}[(Z^a_{-j}\delta^{(2)}_\eta)^2]^{1/2}\\
&\leqslant \bar B_{1n}(|\delta_\theta|\vee\|\eta-\eta_{uj}\|_e)
\end{aligned}$$

where $\bar B_{1n}\leqslant C$ by the same arguments of bounded (second) moments of linear combinations.

Assumption 1(v)(c) follows similarly as

$$\begin{aligned}
&\partial^2_r\mathrm{E}(\psi_{uj}(W_{uj},\theta+r\delta_\theta,\eta+r\delta_\eta))\big|_{r=\bar r}=\\
&-\mathrm{E}[K_\varpi(W)f'_{X_a|X_{-a},\varpi}(t_{\bar r})(Z^a_j-Z^a_{-j}\eta^{(2)}_{\bar r})\{Z^a_j\delta_\theta+Z^a_{-j}\delta^{(1)}_\eta+\delta^{(3)}_\eta\}^2]\\
&+2\mathrm{E}[K_\varpi(W)f_{X_a|X_{-a},\varpi}(t_{\bar r})(Z^a_{-j}\delta^{(2)}_\eta)\{Z^a_j\delta_\theta+Z^a_{-j}\delta^{(1)}_\eta+\delta^{(3)}_\eta\}]
\end{aligned}$$

and under $|f'_{X_a|X_{-a},\varpi}|\leqslant\bar f'_n$, from Cauchy-Schwartz inequality we have

$$\begin{aligned}
&\big|\partial^2_r\mathrm{E}(\psi_{uj}(W_{uj},\theta+r\delta_\theta,\eta+r\delta_\eta))\big|_{r=\bar r}\big|\\
&\leqslant|\bar f'_n\mathrm{E}[(Z^a_j-Z^a_{-j}\eta^{(2)}_{\bar r})^2]^{1/2}\{\mathrm{E}[(Z^a_j)^4]|\delta_\theta|^2+\mathrm{E}[(Z^a_{-j}\delta^{(1)}_\eta)^4]^{1/2}\}+C\mathrm{E}[\{\delta^{(3)}_\eta\}^2]\\
&+2\bar f\mathrm{E}[(Z^a_{-j}\delta^{(2)}_\eta)^2]^{1/2}\{\mathrm{E}[(Z^a_j)^2]^{1/2}|\delta_\theta|+\mathrm{E}[(Z^a_{-j}\delta^{(1)}_\eta)^2]^{1/2}+\mathrm{E}[\{\delta^{(3)}_\eta\}^2]^{1/2}\}\\
&\leqslant\bar B_{2n}(\delta^2_\theta\vee\|\eta-\eta_{uj}\|^2_e)
\end{aligned}$$

where $\bar B_{2n}\leqslant C(1+\bar f'_n)$ by the same arguments of bounded (fourth) moments as before and using that $|\mathrm{E}[(Z^a_j-Z^a_{-j}\eta^{(2)}_{\bar r})\{\delta^{(3)}_\eta\}^2]|\leqslant\{\mathrm{E}[(Z^a_j-Z^a_{-j}\eta^{(2)}_{\bar r})^2\{\delta^{(3)}_\eta\}^2\}^{1/2}\mathrm{E}[\{\delta^{(3)}_\eta\}^2]^{1/2}\leqslant C\mathrm{E}[\{\delta^{(3)}_\eta\}^2]$.

To verify the near orthogonality condition, note that for all $u\in\mathcal{U}$ and $j\in[p]$, since by definition $f_u=f_{X_a|X_{-a},\varpi}(Z^a\beta_u+r_u)$ we have

$$|\mathrm{D}_{u,j,0}[\tilde\eta_{uj}-\eta_{uj}]|=|-\mathrm{E}[K_\varpi(W)f_u\{Z^a_{-j}(\tilde\eta^{(2)}-\eta^{(2)}_{uj})+r_u\}v_{uj}]|\leqslant\delta_n n^{-1/2}$$

by the relations $\mathrm{E}[K_\varpi(W)(\tau-F_{X_a|X_{-a},\varpi}(Z^a\beta_u+r_u))Z^a_{-j}]=0$ and $\mathrm{E}[K_\varpi(W)f_uZ^a_{-j}v_{uj}]=0$ implied by the model, and $|\mathrm{E}[K_\varpi(W)f_ur_uv_{uj}]|\leqslant\delta_n n^{-1/2}$ by Condition CI. Thus, condition (H.58) holds.

Furthermore, since $\Theta_{uj} \subset \theta_{uj} \pm C/\log n$, for $J_{uj} = \partial_\theta \mathrm{E}[\psi_{uj}(W_{uj}, \theta_{uj}, \eta_{uj})] = \mathrm{E}[K_\varpi(W) f_u Z^a_j v_{uj}] = \mathrm{E}[K_\varpi(W) v^2_{uj}] = \mathrm{E}[v^2_{uj} \mid \varpi]\mathrm{P}(\varpi)$ as $\mathrm{E}[K_\varpi(W) f_u Z^a_{-j} v_{uj}] = 0$, we have that for all $\theta \in \Theta_{uj}$

$$\mathrm{E}[\psi_{uj}(W_{uj}, \theta, \eta_{uj})] = J_{uj}(\theta - \theta_{uj}) + \frac{1}{2}\partial^2_\theta \mathrm{E}[\psi_{uj}(W_{uj}, \bar\theta, \eta_{uj})](\theta - \theta_{uj})^2$$

where $|\partial^2_\theta \mathrm{E}[\psi_{uj}(W_{uj}, \bar\theta, \eta_{uj})]| \leqslant \bar f'_n \mathrm{E}[|Z^a_j|^2 |v_{uj}| \mid \varpi] P(\varpi) \leqslant \bar f' \mathrm{E}[|Z^a_j|^4 \mid \varpi]^{1/2} \mathrm{E}[|v_{uj}|^2 \mid \varpi]^{1/2} P(\varpi) \leqslant C \bar f' P(\varpi)$ so that for all $\theta \in \Theta_{uj}$

$$|\mathrm{E}[\psi_{uj}(W_{uj}, \theta, \eta_{uj})]| \geqslant \{|\mathrm{E}[v^2_{uj} \mid \varpi]| - \{C^2 \bar f'\}/\log n\}\mathrm{P}(\varpi)|\theta - \theta_{uj}|$$

and we can take $j_n \geqslant c \inf_{\varpi \in \mathcal{W}} \mathrm{P}(\varpi) = c\mu_{\mathcal{W}}$.

Next we verify Assumption 2 with $\mathcal{H}_{ujn} = \{\eta = (0, \beta, \gamma) : \|\beta\|_0 \leqslant Cs, \|\gamma\|_0 \leqslant Cs, \|\beta - \beta_{u,-j}\| \leqslant C\tau_n, \|\gamma - \gamma^j_u\| \leqslant C\tau_n, \|\gamma - \gamma^j_u\|_1 \leqslant C\sqrt{s}\tau_n\}$. We will show that $\widehat\eta_{uj} = (\widetilde\beta_{u,-j}, \widetilde\gamma^j_u, 0) \in \mathcal{H}_{ujn}$ with probability $1 - o(1)$, uniformly over $u \in \mathcal{U}$ and $j \in [p]$.

Under Condition CI and the choice of penalty parameters, by Theorems 1 and 2, with probability $1 - o(1)$, uniformly over $u \in \mathcal{U}$ we have

$$\|\widetilde\beta_u - \beta_u\| \leqslant C\tau_n, \quad \max_{j \in [p]} \sup_{u \in \mathcal{U}} \|\widetilde\gamma^j_u - \gamma^j_u\| \leqslant C\tau_n, \quad \text{and} \quad \max_{j \in [p]} \sup_{u \in \mathcal{U}} \|\widetilde\gamma^j_u\|_0 \leqslant \bar C s,$$

further by thresholding we can achieve $\sup_{u \in \mathcal{U}} \|\widetilde\beta_u\|_0 \leqslant \bar C s$ using Lemma 16.

Next we establish the entropy bounds. For $\eta \in \mathcal{H}_{ujn}$ we have that

$$\psi_{uj}(W_{uj}, \theta, \eta) = K_\varpi(W)(\tau - 1\{X_a \leqslant Z^a_j \theta + Z^a_{-j}\beta_{-j}\})\{Z^a_j - Z^a_{-j}\gamma\}$$

It follows that $\mathcal{F}_1 \subset \mathcal{W}\mathcal{G}_1\mathcal{G}_2 \cup \bar{\mathcal{F}}_0$ where $\bar{\mathcal{F}}_0 = \{\psi_{uj}(W_{uj}, \theta, \eta_{uj}) : u \in \mathcal{U}, j \in [p], \theta \in \Theta_{uj}\}$, $\mathcal{G}_1 = \{\tau - 1\{X_a \leqslant Z^a\beta\} : \|\beta\|_0 \leqslant Cs, \tau \in \mathcal{T}, a \in V\}$, $\mathcal{G}_2 = \{Z^a \to Z^a(1, -\gamma), \|\gamma\|_0 \leqslant Cs, \|\gamma\| \leqslant C, a \in V\}$. By Assumption $\mathcal{W}$ is a VC class of sets with VC index d_W (fixed). It follows that $\mathcal{G}_2$ and $\mathcal{G}_3$ are p choose $O(s)$ VC-subgraph classes with VC indices at most $O(s)$. Therefore, $\mathrm{ent}(\mathcal{G}_1) \vee \mathrm{ent}(\mathcal{G}_2) \vee \mathrm{ent}(\mathcal{W}) \leqslant Cs\log(a_n/\varepsilon) + Cd_W \log(e/\varepsilon)$ by Theorem 2.6.7 in [61] and by standard arguments. Moreover, an envelope F_G for $\mathcal{F}_1$ satisfies

$$\begin{aligned}
\mathrm{E}[F^q_G] &= \mathrm{E}[\sup_{u \in \mathcal{U}, j \in [p], \|\gamma - \gamma^j_u\|_1 \leqslant C\sqrt{s}\tau_n} |v_{uj} - Z^a_{-j}(\gamma - \gamma^j_u)|^q] \\
&\leqslant 2^{q-1}\mathrm{E}[\sup_{u \in \mathcal{U}, j \in [p]} |v_{uj}|^q] + 2^{q-1}\mathrm{E}[\max_{a \in V} \|Z^a\|^q_\infty]\{C\sqrt{s}\tau_n\}^q \\
&\leqslant 2^{q-1}L^q_n + 2^{q-1}\{M_n C\sqrt{s}\tau_n\}^q \leqslant 2^q L^q_n
\end{aligned}$$

since $M_n C\sqrt{s}\tau_n \leqslant \delta_n L_n$ and $\delta_n \leqslant 1$ for n large.

Next we bound the entropy in $\bar{\mathcal{F}}_0$. Note that for any $\psi_{uj}(W_{uj}, \theta, \eta_{uj}) \in \bar{\mathcal{F}}_0$, there is some $\delta \in [-C, C]$ such that

$$\psi_{uj}(W_{uj}, \theta, \eta_{uj}) = K_\varpi(W)\{\tau - 1\{X_a \leqslant Z^a_j\delta + Q_{X_a}(\tau \mid X_{-a}, \varpi)\}\}v_{uj}$$

and therefore $\bar{\mathcal{F}}_0 \subset \mathcal{W}\{\mathcal{T} - \phi(\mathcal{V})\}\mathcal{L}$ where $\phi(t) = 1\{t \leqslant 0\}$, $\mathcal{V} = \cup_{a \in V, j \in [p]} \mathcal{V}_{aj}$ with

$$\mathcal{V}_{aj} := \{X_a - Z^a_j\delta - Q_{X_a}(\tau \mid X_{-a}, \varpi) : \tau \in \mathcal{T}, \varpi \in \mathcal{W}, |\delta| \leqslant C\},$$

and $\mathcal{L} = \cup_{a \in V, j \in [p]}(\mathcal{L}_{aj} + \{v_{\bar u j}\})$ where $\mathcal{L}_{aj} = \{(X, W) \mapsto v_{uj} - v_{\bar u j} = Z^a_{-j}(\gamma^j_u - \gamma^j_{\bar u}) : u \in \mathcal{U}\}$. Note that each $\mathcal{V}_{aj}$ is a VC subgraph class of functions with index $1 + Cd_W$ as $\{Q_{X_a}(\tau \mid X_{-a}, \varpi) : (\tau, \varpi) \in \mathcal{W} \times \mathcal{T}\}$

is a VC-subgraph with VC-dimension Cd_W for every $a \in V$. Since ϕ is monotone, $\phi(\mathcal{V})$ is also the union of VC-dimension of order $1 + Cd_W$.

Letting $F_1 = 1$ as an envelope for $\mathcal{W}$ and $\mathcal{T} - \phi(\mathcal{V})$. By Lemma 5, it follows that $\|\gamma_u^j - \gamma_{u'}^j\| \leqslant L_\gamma\{\|u-u'\|+\|\varpi-\varpi'\|^{1/2}\}$ for some L_γ satisfying $\log(L_\gamma) \leqslant C\log(p|V|n)$ under Condition CI. Therefore, $|v_{uj} - v_{\bar{u}j}| = |Z^a_{-j}(\gamma_u^j - \gamma_{\bar{u}}^j)| \leqslant \|Z^a\|_\infty\sqrt{p}\|\gamma_u^j - \gamma_{\bar{u}}^j\|$. For a choice of envelope $F_a = M_n^{-1}\|Z^a\|_\infty + 2\sup_{u\in\mathcal{U}}|v_{uj}|$ which satisfies $\|F_a\|_{P,q} \lesssim L_n$, we have

$$\begin{aligned}\log N(\epsilon\|F_a\|_{Q,2}, \mathcal{L}_{aj}, \|\cdot\|_{Q,2}) &\leqslant \log N(\tfrac{\epsilon}{M_n}\|\, \|Z^a\|_\infty\|_{Q,2}, \mathcal{L}_{aj}, \|\cdot\|_{Q,2})\\ &\leqslant \log N(\epsilon/\{M_n\sqrt{p}L_\gamma\}, \mathcal{U}, |\cdot|) \leqslant Cd_u\log(M_n pL_\gamma/\epsilon)\end{aligned}$$

Since $\mathcal{L} = \cup_{a\in V, j\in[p]}(\mathcal{L}_j + \{v_{\bar{u}j}\})$, taking $F_L = \max_{a\in V} F_a$, we have that

$$\begin{aligned}\log N(\epsilon\|F_LF_1\|_{Q,2}, \mathcal{F}_0, \|\cdot\|_{Q,2}) &\leqslant \log N(\tfrac{\epsilon}{4}\|F_1\|_{Q,2}, \mathcal{W}, \|\cdot\|_{Q,2}) + \log N(\tfrac{\epsilon}{4}\|F_1\|_{Q,2}, \mathcal{T}-\phi(\mathcal{V}), \|\cdot\|_{Q,2})\\ &\quad + \log\textstyle\sum_{a\in V, j\in[p]} N(\tfrac{\epsilon}{2}\|F_a\|_{Q,2}, \mathcal{L}_{aj}, \|\cdot\|_{Q,2})\\ &\leqslant \log(p|V|) + 1 + C'\{d_W + d_u\}\log(4eM_n|V|pL_\gamma/\epsilon)\end{aligned}$$

where the last line follows from the previous bounds.

Next we verify the growth conditions in Assumption 2 with the proposed $\mathcal{F}_1$ and $K_n \lesssim CL_n$. We take $s_{n(\mathcal{U},p)} = (1+d_W)s$ and $a_n = \max\{n, p, |V|\}$. Recall that $\bar{B}_{1n} \leqslant C$, $\bar{B}_{2n} \leqslant C$, $j_n \geqslant c\mu_{\mathcal{W}}$. Thus, we have $\sqrt{n}(\tau_n/j_n)^2 \lesssim \sqrt{n}\frac{s\log(p|V|n)}{n(1\wedge \underline{f}_{\mathcal{U}}^2)\mu_{\mathcal{W}}^3} \leqslant \delta_n$ under $s^2\log^2(p|V|n) \leqslant n(1\wedge \underline{f}_{\mathcal{U}}^4)\mu_{\mathcal{W}}^6$. Moreover, $(\tau_n/j_n)^{\alpha/2}\sqrt{s_{n(\mathcal{U},p)}\log(a_n)} \lesssim \sqrt[4]{\frac{(1+d_W)^3s^3\log^3(p|V|n)}{n(1\wedge\underline{f}_{\mathcal{U}}^2)\mu_{\mathcal{W}}^3}} \lesssim \delta_n$ under d_W fixed and $s^3\log^3(p|V|n) \leqslant \delta_n^4 n(1\wedge \underline{f}_{\mathcal{U}}^2)\mu_{\mathcal{W}}^3$ and $s_{n(\mathcal{U},p)}n^{\frac{1}{q}-\frac{1}{2}}K_n\log(a_n)\log n \lesssim (1+d_W)sn^{\frac{1}{q}-\frac{1}{2}}M_n\log(p|V|n)\log n \leqslant \delta_n$ under our conditions. Finally, the conditions of Corollary 4 hold with $\rho_n = (1+d_W)$ since the score is the product of VC-subgraph classes of function with VC index bounded by $C(1+d_W)$. ∎

Proof of Theorem 4. We will invoke Lemma 7 with $\bar{\beta}_u$ as the estimand and $r_{ui} = X_{-a}(\beta_u - \bar{\beta}_u)$, therefore $\mathrm{E}[(\tau - 1\{X_a \leqslant X_{-a}\bar{\beta} + r_u\})X_{-a}] = 0$. To invoke the lemma we verify that the events Ω_1, Ω_2, Ω_3 and Ω_4 hold with probability $1 - o(1)$

$$\begin{aligned}\Omega_1 &:= \{\lambda_u \geqslant c|S_{uj}|/\{\mathbb{E}_n[K_w(W)X_j^2]\}^{1/2}, \text{ for all } u\in\mathcal{U}, j\in V\},\\ \Omega_2 &:= \{\widehat{R}_u(\bar{\beta}_u) \leqslant \bar{R}_{u\gamma} : u\in\mathcal{U}\}\\ \Omega_3 &:= \left\{\sup_{u\in\mathcal{U}, 1/\sqrt{n}\leqslant\|\delta\|_{1,\varpi}\leqslant\sqrt{n}}|\mathbb{E}_n[g_u(\delta,X,W) - \mathrm{E}[g_u(\delta,X,W)\mid X_{-a},W]]|/\|\delta\|_{1,\varpi} \leqslant t_3\right\}\\ \Omega_4 &:= \{K_u\{\mathbb{E}_n[K_w(W)X_j^2]\}^{1/2} \geqslant |\mathbb{E}_n[h_{uj}(X_{-a},W)]|, \text{for all } u\in\mathcal{U}, j\in V\setminus\{a\}\}\end{aligned}$$

where $h_{uj}(X_{-a},W) := \mathrm{E}[K_\varpi(W)\{\tau - F_{X_a|X_{-a},W}(X_{-a}\bar{\beta}_u + r_u)\}X_j \mid X_{-a},W]$.

By Lemma 8 with $\xi = 1/n$, by setting $\lambda_u = c\lambda_0 = c2(1+1/16)n^{-1/2}\sqrt{2\log(8|V|^2\{ne/d_W\}^{2d_W}n)}$, we have $\mathrm{P}(\Omega_1) = 1 - o(1)$. By Lemma 2, setting $R_{u\gamma} = Cs(1+d_W)\log(|V|n)/n$ we have $\mathrm{P}(\Omega_2) = 1 - o(1)$ for some $\gamma = o(1)$. By Lemma 10 we have $\mathrm{P}(\Omega_3) = 1 - o(1)$ by setting $t_3 := C\sqrt{(1+d_W)\log(|V|nM_n/\gamma)}$. Finally, by Lemma 11 with $K_u = C\sqrt{\frac{(1+d_W)\log(|V|n)}{n}}$ we have $\mathrm{P}(\Omega_4) = 1 - o(1)$

It follows that with probability $1-o(1)$ that $\|\beta_u\|_{1,\varpi} \leqslant \|\beta_u\| \leqslant \sqrt{s}C \frac{1}{\lambda_u(1-1/c)}\bar{R}_{u\gamma} \leqslant \sqrt{n}$ for all $u \in \mathcal{U}$, and (F.44) holds for all $\delta \in A_u := \Delta_{\varpi,2\mathbf{c}} \cup \{v : \|v\|_{1,\varpi} \leqslant 2\mathbf{c}\bar{R}_{u\gamma}/\lambda_u\}$, $q_{A_u}/4 \geqslant (\sqrt{\bar{f}}+1)\|r_u\|_{n,\varpi} + [\lambda_u + t_3 + K_u]\frac{3\mathbf{c}\sqrt{s}}{\kappa_{u,2\mathbf{c}}}$ and $q_{A_u} \geqslant \{2\mathbf{c}\left(1+\frac{t_3+K_u}{\lambda_u}\right)\bar{R}_{u\gamma}\}^{1/2}$. Then

By Lemma 7, we have uniformly over all $u = (a,\tau,\varpi) \in \mathcal{U} := V \times \mathcal{T} \times \mathcal{W}$

$$\|\sqrt{f_u}X_{-a}(\widehat{\beta}_u - \beta_u)\|_{n,\varpi} \leqslant C\sqrt{\frac{(1+d_W)\log(n|V|)}{n}}\frac{\sqrt{s}}{\kappa_{u,2\mathbf{c}}} \quad \text{and} \quad \|\widehat{\beta}_u - \beta_u\|_{1,\varpi} \leqslant C\sqrt{\frac{(1+d_W)\log(n|V|)}{n}}\frac{s}{\kappa_{u,2\mathbf{c}}}$$

where $\kappa_{u,2\mathbf{c}}$ is bounded away from zero with probability $1-o(1)$ for n sufficiently large. Consider the thresholded estimators $\widehat{\beta}_u^\mu$ for $\mu = \{(1+d_W)\log(n|V|)/n\}^{1/2}$. By Lemma 16 we have $\|\widehat{\beta}_u^\mu\|_0 \leqslant Cs$ and the same rates of convergence as $\widehat{\beta}_u$. Therefore, by refitting over the support of $\widehat{\beta}_u^\mu$ we have by Lemma 14, the estimator $\widetilde{\beta}_u$ has the same rate of convergence where we have that $\widehat{Q}_u \lesssim Cs(1+d_W)\log(|V|n)/n$.

Next we will invoke Lemma 7 for the new penalty choice and penalty loadings. (We note that minor modifications cover the new penalty loadings.)

$$\begin{aligned}
\Omega_1 &:= \{\lambda_u \geqslant c|S_{uj}|/\{\mathbb{E}_n[K_w(W)\epsilon_u^2 X_j^2]\}^{1/2}, \text{ for all } u \in \mathcal{U}, j \in V\},\\
\Omega_2 &:= \{\widehat{R}_u(\bar{\beta}_u) \leqslant \bar{R}_{u\gamma} : u \in \mathcal{U}\}\\
\Omega_3 &:= \left\{\sup\nolimits_{u\in\mathcal{U},1/\sqrt{n}\leqslant\|\delta\|_{1,\varpi}\leqslant\sqrt{n}} |\mathbb{E}_n[g_u(\delta,X,W) - \mathrm{E}[g_u(\delta,X,W) \mid X_{-a},W]]|/\{\theta_u\|\delta\|_{1,\varpi}\} \leqslant t_3\right\}\\
\Omega_4 &:= \{K_u\theta_u\{\mathbb{E}_n[K_w(W)X_j^2]\}^{1/2} \geqslant |\mathbb{E}_n[h_{uj}(X_{-a},W)]|, \text{for all } u \in \mathcal{U}, j \in V\setminus\{a\}\}\\
\Omega_5 &:= \{\theta_u \geqslant \max\nolimits_{j\in V}\{\mathbb{E}_n[K_\varpi(W)X_j^2]/\mathbb{E}_n[K_\varpi(W)\epsilon_u^2X_j^2]\}^{1/2}\}
\end{aligned}$$

where event Ω_5 simply makes the relevant norms equivalent, $\|\cdot\|_{1,u} \leqslant \|\cdot\|_{1,\varpi} \leqslant \theta_u\|\cdot\|_{1,u}$. Note that we can always take $\theta_u \leqslant 1/\{\tau(1-\tau)\} \leqslant C$ since $\mathcal{T}$ is a fixed compact set.

Next we show that the bootstrap approximation of the score provides a valid choice of penalty parameter. Let $\widehat{\epsilon}_u := 1\{X_a \leqslant X_{-a}\widetilde{\beta}_u\} - \tau$. For notational convenience for $u \in \mathcal{U}$, $j \in V\setminus\{a\}$ define

$$\widehat{\psi}_{uj,i} := \frac{K_\varpi(W_i)\widehat{\epsilon}_{ui}X_{ij}}{\mathbb{E}_n[K_\varpi(W)\widehat{\epsilon}_{a\tau\varpi}^2X_j^2]^{1/2}}, \quad \bar{\psi}_{uj,i} = \frac{K_\varpi(W_i)\epsilon_{ui}X_{ij}}{\mathrm{E}[K_\varpi(W)\epsilon_{a\tau\varpi}^2X_j^2]^{1/2}}, \quad \psi_{uj,i} = \frac{K_\varpi(W_i)\epsilon_{ui}X_{ij}}{\mathbb{E}_n[K_\varpi(W)\epsilon_{a\tau\varpi}^2X_j^2]^{1/2}}$$

We will consider the following processes:

$$\widehat{\mathcal{G}}_{uj} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i\widehat{\psi}_{uj,i} \quad \overline{\mathcal{G}}_{uj} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i\bar{\psi}_{uj,i} \quad \bar{S}_{uj} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \bar{\psi}_{uj,i} \quad S_{uj} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_{uj,i}$$

and $\mathcal{N}$ is a tight zero-mean Gaussian process with covariance operator given by $\mathrm{E}[\bar{\psi}_{uj}\bar{\psi}_{u'j'}]$. Their supremum are denoted by $Z_S := \sup_{u\in\mathcal{U},j\in V\setminus\{a\}}|S_{uj}|$, $\bar{Z}_S := \sup_{u\in\mathcal{U},j\in V\setminus\{a\}}|\bar{S}_{uj}|$, $\bar{Z}_G^* := \sup_{u\in\mathcal{U},j\in V\setminus\{a\}}|\mathcal{G}_{uj}|$, $\widehat{Z}_G^* := \sup_{u\in\mathcal{U},j\in V\setminus\{a\}}|\widehat{\mathcal{G}}_{uj}|$, and $Z_N := \sup_{u\in\mathcal{U},j\in V\setminus\{a\}}|\mathcal{N}_{uj}|$.

The penalty choice should majorate Z_S and we simulate via $\widehat{Z}_G^*$. We have that

$$\begin{aligned}
|P(Z_S \leqslant t) - P(\widehat{Z}_G^* \leqslant t)| &\leqslant |P(Z_S \leqslant t) - P(\bar{Z}_S \leqslant t)| + |P(\bar{Z}_S \leqslant t) - P(Z_N \leqslant t)|\\
&\quad + |P(Z_N \leqslant t) - P(\bar{Z}_G^* \leqslant t)| + |P(\bar{Z}_G^* \leqslant t) - P(\widehat{Z}_G^* \leqslant t)|
\end{aligned}$$

We proceed to bound each term. We have that

$$\begin{aligned}
|Z_S - \bar{Z}_S| &\leqslant \bar{Z}_S \sup_{u\in\mathcal{U},j\in V\setminus\{a\}}\left|\frac{\mathrm{E}[K_\varpi(W)\varepsilon_u^2X_j^2]^{1/2}}{\mathbb{E}_n[K_\varpi(W)\varepsilon_u^2X_j^2]^{1/2}} - 1\right|\\
&\leqslant \bar{Z}_S \sup_{u\in\mathcal{U},j\in V\setminus\{a\}}\left|\frac{(\mathbb{E}_n - \mathrm{E})[K_\varpi(W)\varepsilon_u^2X_j^2]}{\mathbb{E}_n[K_\varpi(W)\varepsilon_u^2X_j^2]^{1/2}\{\mathbb{E}_n[K_\varpi(W)\varepsilon_u^2X_j^2]^{1/2} + \mathrm{E}[K_\varpi(W)\varepsilon_u^2X_j^2]^{1/2}\}}\right|
\end{aligned}$$

Therefore, since $\{1\{X_a \leqslant X_{-a}\beta_u\} : u \in \mathcal{U}\}$ is a VC-subgraph of VC dimension $1 + d_W$, and $\mathcal{W}$ is a VC class of sets of dimension d_W, we apply Lemma 18 with envelope $F = \|X\|_\infty^2$ and $\sigma^2 \leqslant \max_{j\in V} \mathrm{E}[X_j^4] \leqslant C$ to obtain with probability $1 - o(1)$

$$\sup_{u\in\mathcal{U}, j\in V\setminus\{a\}} |(\mathbb{E}_n - \mathrm{E})[K_\varpi(W)\varepsilon_u^2 X_j^2]| \lesssim \delta_{1n} := \sqrt{\frac{(1+d_W)\log(|V|n)}{n}} + \frac{M_n^2(1+d_W)\log(|V|n)}{n}$$

where $\delta_{1n} = o(\mu_{\mathcal{W}}^2)$ under Condition P. Note that this implies that the denominator above is bounded away from zero by $c\mu_{\mathcal{W}}$. Therefore,

$$|Z_S - \bar{Z}_S| \lesssim_P \bar{Z}_S \delta_{1n}/\mu_{\mathcal{W}}.$$

By Theorem 2.1 in [27], since $\mathrm{E}[\bar{\psi}_{uj}^4] \leqslant C$, there is a version of Z_N such that

$$|\bar{Z}_S - Z_N| \lesssim_P \delta_{2n} := \left(\frac{M_n(1+d_W)\log(n|V|)}{n^{1/2}} + \frac{M_n^{1/3}((1+d_W)\log(n|V|))^{2/3}}{n^{1/4}}\right)$$

and there is also a version of

$$|Z_N - \bar{Z}_G^*| \lesssim_P \left(\frac{M_n(1+d_W)\log(n|V|)}{n^{1/2}} + \frac{M_n^{1/3}((1+d_W)\log(n|V|))^{2/3}}{n^{1/4}}\right)$$

Finally, we have that

$$|\bar{Z}_G^* - \widehat{Z}_G^*| \leqslant \sup_{u\in\mathcal{U}, j} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i(\widehat{\psi}_{uj,i} - \bar{\psi}_{uj,i}) \right|$$

where conditional on $(X_i, W_i), i = 1, \ldots, n$, $\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i(\widehat{\psi}_{uj,i} - \bar{\psi}_{uj,i})$ is a zero-mean Gaussian with variance $\mathbb{E}_n[(\widehat{\psi}_{uj,i} - \bar{\psi}_{uj,i})^2] \leqslant \bar{\delta}_n$. Note that $\widehat{\epsilon}_u = 1\{X_a \leqslant X_{-a}\widetilde{\beta}_u\} - \tau$ where $\|\widetilde{\beta}_u\|_0 \leqslant Cs$. Therefore, we have $\{1\{X_a \leqslant X_{-a}\widetilde{\beta}_u\} : u \in \mathcal{U}\} \subset \{1\{X_a \leqslant X_{-a}\beta\} : \|\beta\|_0 \leqslant Cs\}$ which is the union of $\binom{|V|}{Cs}$ VC subgraph classes of functions with VC dimension $C's$. Therefore,

$$\begin{aligned}\sup_{u\in\mathcal{U}, j\in V\setminus\{a\}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i(\widehat{\psi}_{uj,i} - \bar{\psi}_{uj,i}) \right| &\lesssim_P \bar{\delta}_n \sqrt{s(1+d_W)\log(|V|n)} \\ &\lesssim_P \delta_{3n} := \{s\log(|V|n)/n\}^{1/4}\sqrt{s(1+d_W)\log(|V|n)}\end{aligned}$$

where $\bar{\delta}_n \lesssim \{s\log(|V|n)/n\}^{1/4}$ by $\mathbb{E}_n[(\widehat{\psi}_{uj,i} - \bar{\psi}_{uj,i})^2]^{1/2} \leqslant |\mathbb{E}_n[(\widehat{\psi}_{uj,i} - \bar{\psi}_{uj,i})^2 - \mathrm{E}[(\widehat{\psi}_{uj,i} - \bar{\psi}_{uj,i})^2]]^{1/2} + \mathrm{E}[(\widehat{\psi}_{uj,i} - \bar{\psi}_{uj,i})^2]^{1/2} \lesssim \{s\log(|V|n)/n\}^{1/4}$. The rest of the proof follows similarly to Corollary 2.2 in [12] since under Condition P that $r_n := \delta_{1n} + \delta_{2n} + \delta_{3n} = o(\{\mathrm{E}[Z_N]\}^{-1}))$ where $\mathrm{E}[Z_N] \lesssim \{(1 + d_W)\log(|V|n)\}^{-1/2})$. Then we have $\sup_t |\mathrm{P}(Z_S \leqslant t) - \mathrm{P}(\widehat{Z}_G^* \leqslant t)| = o_P(1)$ which in turn implies that

$$\begin{aligned}\mathrm{P}(\Omega_1) &= \mathrm{P}(Z_S \leqslant \widehat{c}_G^*(\alpha)) \\ &\geqslant \mathrm{P}(\widehat{Z}_G^* \leqslant \widehat{c}_G^*(\alpha)) - |\mathrm{P}(Z_S \leqslant \widehat{c}_G^*(\alpha)) - \mathrm{P}(\widehat{Z}_G^* \leqslant \widehat{c}_G^*(\alpha))| \\ &\geqslant 1 - \alpha + o_P(1)\end{aligned}$$

Note that the occurrence of the events Ω_2, Ω_3 and Ω_4 follows by similar arguments. The result follows by Lemma 7, thresholding and applying Lemma 16 and Lemma 14 similarly to before.

■

Appendix E. Technical Lemmas for Conditional Independence Quantile Graph Model

Let $u = (a, \tau, \varpi) \in \mathcal{U} := V \times \mathcal{T} \times \mathcal{W}$, and $T_u = \text{support}(\beta_u)$ where $|T_u| \leqslant s$ for all $u \in \mathcal{U}$.

Define the pseudo-norms

$$\|v\|_{n,\varpi}^2 := \frac{1}{n}\sum_{i=1}^n K_\varpi(W_i)(v_i)^2, \quad \|\delta\|_{2,\varpi} := \left\{\sum_{j=1}^p \widehat{\sigma}_{a\varpi j}^2 |\delta_j|^2\right\}^{1/2}, \quad \text{and} \quad \|\delta\|_{1,\varpi} := \sum_{j=1}^p \widehat{\sigma}_{a\varpi j}|\delta_j|,$$

where $\widehat{\sigma}_{a\varpi j}^2 = \mathbb{E}_n[\{K_\varpi(W)Z_j^a\}^2]$. These pseudo-norms induce the following restricted eigenvalue as

$$\kappa_{u,\mathbf{c}} = \min_{\|\delta_{T_u^c}\|_{1,\varpi} \leqslant \mathbf{c}\|\delta_{T_u}\|_{1,\varpi}} \frac{\|\sqrt{f_u} Z^a \delta\|_{n,\varpi}}{\|\delta\|_{1,\varpi}/\sqrt{s}}.$$

The restricted eigenvalue $\kappa_{u,c}$ is an counterpart of the restricted eigenvalue proposed in [22] for our setting. We note that $\kappa_{u,c}$ typically will vary with the events $\varpi \in \mathcal{W}$.

Let $u = (a, \tau, \varpi) \in V \times \mathcal{T} \times \mathcal{W} =: \mathcal{U}$. We will consider three key events in our analysis. Let

$$\Omega_1 := \{\lambda_u \geqslant c|S_{uj}|/\widehat{\sigma}_{uj}, \text{ for all } u \in \mathcal{U}, j \in [p]\} \tag{E.33}$$

which occurs with probability at least $1-\gamma$ by the choice the choice of λ_u. (In the case of conditional independence we have $S_u := \mathbb{E}_n[K_\varpi(W)(\tau - 1\{X_a \leqslant Z^a\beta_u + r_u\})Z_j^a]$, $\widehat{\sigma}_{uj}^2 = \mathbb{E}_n[K_\varpi(W)(Z_j^a)^2]$ and $\lambda_u = \lambda_{V\mathcal{T}\mathcal{W}}\sqrt{\tau(1-\tau)}$. In the case of predictive we have $S_u := \mathbb{E}_n[K_\varpi(W)(\tau - 1\{X_a \leqslant X_{-a}'\beta_u\})X_{-a}]$, $\widehat{\sigma}_{uj}^2 = \mathbb{E}_n[K_\varpi(W)X_j^2]$ and $\lambda_u = \lambda_0$.)

To define the next event consider

$$\widehat{R}_u(\beta) = \mathbb{E}_n[K_\varpi(W)\{\rho_u(X_a - Z^a\beta) - \rho_u(X_a - Z^a\beta_u - r_u) - (\tau - 1\{X_a \leqslant Z^a\beta_u + r_u\})(Z^a\beta - Z^a\beta_u - r_u)\}]$$

in the case of the conditional independence. (In the case of predictive quantile graph models we replace Z^a with X_{-a}.) By convexity we have $\widehat{R}_u(\beta) \geqslant 0$. The event

$$\Omega_2 := \{\widehat{R}_u(\beta_u) \leqslant \bar{R}_{u\gamma} : u \in \mathcal{U}\} \tag{E.34}$$

where $\bar{R}_{u\gamma}$ are chosen so that Ω_2 occurs with probability at least $1-\gamma$. Note that by Lemma 2, we have $\mathbb{E}_n\mathrm{E}[\widehat{R}_u(\beta_u) \mid X_{-a}, W] \leqslant \bar{f}\|r_u\|_{n,\varpi}^2/2$ and with probability at least $1-\gamma$, $\widehat{R}_u(\beta_u) \leqslant \bar{R}_{u\gamma} := 4\max\{\bar{f}\|r_u\|_{n,\varpi}^2, \ \|r_u\|_{n,\varpi}C\sqrt{\log(n^{1+d_W}p/\gamma)/n}\} \leqslant C's\log(n^{1+d_W}p/\gamma)/n$.

Define $g_u(\delta, X, W) = K_\varpi(W)\{\rho_\tau(X_a - Z^a(\beta_u + \delta)) - \rho_\tau(X_a - Z^a\beta_u)\}$ so that event Ω_3 is defined as

$$\Omega_3 := \left\{\sup_{u\in\mathcal{U}, 1/\sqrt{n} \leqslant \|\delta\|_{1,\varpi} \leqslant \sqrt{n}} \frac{|\mathbb{E}_n[g_u(\delta, X, W) - \mathrm{E}[g_u(\delta, X, W) \mid X_{-a}, W]]|}{\|\delta\|_{1,\varpi}} \leqslant t_3\right\} \tag{E.35}$$

where t_3 is given in Lemma 3 so that Ω_3 holds with probability at least $1-\gamma$.

Lemma 1. *Suppose that Ω_1, Ω_2 and Ω_3 holds. Further assume $2\frac{1+1/c}{1-1/c}\|\beta_u\|_{1,\varpi} + \frac{1}{\lambda_u(1-1/c)}\bar{R}_{u\gamma} \leqslant \sqrt{n}$ for all $u \in \mathcal{U}$, and (F.44) holds for all $\delta \in A_u := \Delta_{\varpi,2\mathbf{c}} \cup \{v : \|v\|_{1,\varpi} \leqslant 2\mathbf{c}\bar{R}_{u\gamma}/\lambda_u\}$, $q_{A_u}/4 \geqslant (\sqrt{\bar{f}} +$*

$1)\|r_u\|_{n,\varpi} + [\lambda_u + t_3]\,\frac{3\mathbf{c}\sqrt{s}}{\kappa_{u,2\mathbf{c}}}$ *and* $q_{A_u} \geqslant \{2\mathbf{c}\left(1+\frac{t_3}{\lambda_u}\right)\bar{R}_{u\gamma}\}^{1/2}$. *Then uniformly over all* $u = (a,\tau,\varpi) \in \mathcal{U} := V \times \mathcal{T} \times \mathcal{W}$ *we have*

$$\begin{aligned} \|\sqrt{f_u}Z^a(\widehat{\beta}_u - \beta_u)\|_{n,\varpi} &\leqslant \sqrt{8\mathbf{c}\left(1+\tfrac{t_3}{\lambda_u}\right)\bar{R}_{u\gamma}} + (\bar{f}^{1/2}+1)\|r_u\|_{n,\varpi} + \tfrac{3\mathbf{c}\lambda_u\sqrt{s}}{\kappa_{u,2\mathbf{c}}} + t_3\tfrac{(1+\mathbf{c})\sqrt{s}}{\kappa_{u,2\mathbf{c}}} \\ \|\widehat{\beta}_u - \beta_u\|_{1,\varpi} &\leqslant (1+2\mathbf{c})\sqrt{s}\|\sqrt{f_u}Z^a\delta_u\|_{n,\varpi}/\kappa_{u,2\mathbf{c}} + \tfrac{2\mathbf{c}}{\lambda_u}\bar{R}_{u\gamma} \end{aligned}$$

Proof of Lemma 1. Let $u = (a,\tau,\varpi)$, $\delta_u = \widehat{\beta}_u - \beta_u$. By convexity we have $\widehat{R}_u(\beta) \geqslant 0$, and by definition of $\widehat{\beta}_u$ we have

$$\begin{aligned} &\widehat{R}_u(\widehat{\beta}_u) - \widehat{R}_u(\beta_u) + S_u'\delta_u \\ &= \mathbb{E}_n[K_\varpi(W)\rho_u(X_a - Z^a\widehat{\beta}_u)] - \mathbb{E}_n[K_\varpi(W)\rho_u(X_a - Z^a\beta_u)] \\ &\leqslant \lambda_u\|\beta_u\|_{1,\varpi} - \lambda_u\|\widehat{\beta}_u\|_{1,\varpi} \end{aligned} \tag{E.36}$$

where S_u is defined as in (E.33) so that under Ω_1 we have $\lambda_u > c|S_{uj}|/\widehat{\sigma}_{uj}$.

Under $\Omega_1 \cap \Omega_2$, and since $\widehat{R}_u(\beta) \geqslant 0$, we have

$$\begin{aligned} -\widehat{R}_u(\beta_u) - \tfrac{\lambda_u}{c}\|\delta_u\|_{1,\varpi} &\leqslant \widehat{R}_u(\beta_u+\delta_u) - \widehat{R}_u(\beta_u) + \mathbb{E}_n[K_\varpi(W)(\tau - 1\{X_a \leqslant Z^a\beta_u + r_u\})Z^a\delta_u] \\ &= \mathbb{E}_n[K_\varpi(W)\rho_u(X_a - Z^a(\delta_u+\beta_u))] - \mathbb{E}_n[K_\varpi(W)\rho_u(X_a - Z^a\beta_u)] \\ &\leqslant \lambda_u\|\beta_u\|_{1,\varpi} - \lambda_u\|\delta_u+\beta_u\|_{1,\varpi} \end{aligned} \tag{E.37}$$

so that for $\mathbf{c} = (c+1)/(c-1)$

$$\|\delta_{T_u^c}\|_{1,\varpi} \leqslant \mathbf{c}\|\delta_{T_u}\|_{1,\varpi} + \frac{c}{\lambda_u(c-1)}\widehat{R}_u(\beta_u).$$

To establish that $\delta_u \in A_u := \Delta_{\varpi,2\mathbf{c}} \cup \{v : \|v\|_{1,\varpi} \leqslant 2\mathbf{c}\bar{R}_{u\gamma}/\lambda_u\}$ we consider two cases. If $\|\delta_{u,T_u^c}\|_{1,\varpi} \geqslant 2\mathbf{c}\|\delta_{u,T_u}\|_{1,\varpi}$ we have

$$\frac{1}{2}\|\delta_{u,T_u^c}\|_{1,\varpi} \leqslant \frac{c}{\lambda_u(c-1)}\widehat{R}_u(\beta_u)$$

and consequentially

$$\|\delta_u\|_{1,\varpi} \leqslant \{1+1/(2c)\}\|\delta_{u,T_u^c}\|_{1,\varpi} \leqslant \frac{2\mathbf{c}}{\lambda_u}\widehat{R}_u(\beta_u).$$

Otherwise $\|\delta_{u,T_u^c}\|_{1,\varpi} \leqslant 2\mathbf{c}\|\delta_{u,T_u}\|_{1,\varpi}$, and we have

$$\|\delta_u\|_{1,\varpi} \leqslant (1+2\mathbf{c})\|\delta_{u,T_u}\|_{1,\varpi} \leqslant (1+2\mathbf{c})\sqrt{s}\|\sqrt{f_u}Z^a\delta_u\|_{n,\varpi}/\kappa_{u,2\mathbf{c}}.$$

Thus we have $\delta_u \in A_u$ under $\Omega_1 \cap \Omega_2$.

Furthermore, (E.37) also implies that

$$\begin{aligned} \|\delta_u+\beta_u\|_{1,\varpi} &\leqslant \|\beta_u\|_{1,\varpi} + \tfrac{1}{c}\|\delta_u\|_{1,\varpi} + \widehat{R}_u(\beta_u)/\lambda_u \\ &\leqslant (1+1/c)\|\beta_u\|_{1,\varpi} + (1/c)\|\delta_u+\beta_u\|_{1,\varpi} + \widehat{R}_u(\beta_u)/\lambda_u. \end{aligned}$$

which in turn establishes

$$\|\delta_u\|_{1,\varpi} \leqslant 2\frac{1+1/c}{1-1/c}\|\beta_u\|_{1,\varpi} + \frac{1}{\lambda_u(1-1/c)}\widehat{R}_u(\beta_u) \leqslant 2\frac{1+1/c}{1-1/c}\|\beta_u\|_{1,\varpi} + \frac{1}{\lambda_u(1-1/c)}\bar{R}_{u\gamma}$$

where the last inequality holds under Ω_2. Thus, $\|\delta_u\|_{1,\varpi} \leqslant \sqrt{n}$ under our condition. In turn, δ_u is considered in Ω_3.

Under $\Omega_1 \cap \Omega_2 \cap \Omega_3$ we have

$$\begin{array}{l} \mathbb{E}_n \mathrm{E}[K_\varpi(W)\{\rho_u(X_a - Z^a(\beta_u+\delta_u)) - \rho_u(X_a - Z^a\beta_u)\} \mid X_{-a}, W] \\ \leqslant \mathbb{E}_n[K_\varpi(W)\{\rho_u(X_a - Z^a(\beta_u+\delta_u)) - \rho_u(X_a - Z^a\beta_u)\}] + \|\delta_u\|_{1,\varpi} t_3 \\ \leqslant \lambda_u \|\delta_u\|_{1,\varpi} + \|\delta_u\|_{1,\varpi} t_3 \\ \leqslant 2\mathbf{c}\left(1+\frac{1}{\lambda_u}t_3\right)\bar{R}_{u\gamma} + \|\sqrt{f_u}Z^a\delta_u\|_{n,\varpi}\,[\lambda_u + t_3]\,\frac{3\mathbf{c}\sqrt{s}}{\kappa_{u,2\mathbf{c}}} \end{array} \tag{E.38}$$

where we used the bound $\|\delta_u\|_{1,\varpi} \leqslant (1+2\mathbf{c})\sqrt{s}\|\sqrt{f_u}Z^a\delta_u\|_{n,\varpi}/\kappa_{u,2\mathbf{c}} + \frac{2\mathbf{c}}{\lambda_u}\bar{R}_{u\gamma}$ under $\Omega_1 \cap \Omega_2$.

Using Lemma 12, since (F.44) holds, we have for each $u \in \mathcal{U}$

$$\begin{array}{r} \mathbb{E}_n \mathrm{E}[K_\varpi(W)\{\rho_u(X_a - Z^a(\beta_u+\delta_u)) - \rho_u(X_a - Z^a\beta_u)\} \mid X_{-a}, W] \\ \geqslant -(\sqrt{\bar{f}}+1)\|r_u\|_{n,\varpi}\|\sqrt{f_u}Z^a\delta\|_{n,\varpi} - \max_{j\in[p]} |\mathbb{E}_n[\mathrm{E}[S_{uj} \mid X_{-a}, W]/\widehat{\sigma}_{uj}]|\ \|\delta_u\|_{1,\varpi} \\ + \frac{\|\sqrt{f_u}Z^a\delta_u\|_{n,\varpi}^2}{4} \wedge \bar{q}_{A_u}\|\sqrt{f_u}Z^a\delta_u\|_{n,\varpi} \end{array}$$

where we have $\mathrm{E}[S_{i,uj} \mid X_{i,-a}, W_i] = 0$ since $\tau = \mathrm{P}(X_a \leqslant Z^a\beta_u + r_u \mid X_{-a}, W)$ by definition of the conditional quantile.

Note that for positive numbers $(t^2/4) \wedge qt \leqslant A + Bt$ implies $t^2/4 \leqslant A + Bt$ provided $q/2 > B$ and $2q^2 > A$. (Indeed, otherwise $(t^2/4) \geqslant qt$ so that $t \geqslant 4q$ which in turn implies that $2q^2 + qt/2 \leqslant (t^2/4)\wedge qt \leqslant A+Bt$.) Since $q_{A_u}/4 \geqslant (\sqrt{\bar{f}}+1)\|r_u\|_{n,\varpi} + \left[\{\lambda_u + t_3\}\frac{3\mathbf{c}\sqrt{s}}{\kappa_{u,2\mathbf{c}}}\right]$ and $q_{A_u} \geqslant \{2\mathbf{c}\left(1+\frac{t_3}{\lambda_u}\right)\bar{R}_{u\gamma}\}^{1/2}$, the minimum on the right hand side is achieved by the quadratic part for all $u \in \mathcal{U}$. Therefore we have uniformly over $u \in \mathcal{U}$

$$\frac{\|\sqrt{f_u}Z^a\delta_u\|_{n,\varpi}^2}{4} \leqslant 2\mathbf{c}\left(1+\frac{t_3}{\lambda_u}\right)\bar{R}_{u\gamma} + \|\sqrt{f_u}Z^a\delta_u\|_{n,\varpi}\left[(\sqrt{\bar{f}}+1)\|r_u\|_{n,\varpi} + \{\lambda_u+t_3\}\frac{3\mathbf{c}\sqrt{s}}{\kappa_{u,2\mathbf{c}}}\right]$$

which implies that

$$\|\sqrt{f_u}Z^a\delta_u\|_{n,\varpi} \quad \leqslant \sqrt{8\mathbf{c}\left(1+\frac{t_3}{\lambda_u}\right)\bar{R}_{u\gamma}} + \left[(\sqrt{\bar{f}}+1)\|r_u\|_{n,\varpi} + \{\lambda_u+t_3\}\frac{3\mathbf{c}\sqrt{s}}{\kappa_{u,2\mathbf{c}}}\right].$$

■

Lemma 2 (CIQGM, Event Ω_2)**.** *Under Condition CI we have* $\mathbb{E}_n[\mathrm{E}[\widehat{R}_u(\beta_u) \mid X_{-a}, \varpi]] \leqslant \bar{f}\|r_{ui}\|_{n,\varpi}^2/2$, $\widehat{R}_u(\beta_u) \geqslant 0$ *and*

$$\mathrm{P}(\sup_{u\in\mathcal{U}} \widehat{R}_u(\beta_u) \geqslant C\{1+\bar{f}\}\{n^{-1}s(1+d_W)\log(p|V|n)\}) = 1 - o(1).$$

Proof of Lemma 2. We have that $\widehat{R}_u(\beta_u) \geqslant 0$ by convexity of ρ_τ. Let $\epsilon_{ui} = X_{ai} - Z_i^a\beta_u - r_{ui}$ where $\|\beta_u\|_0 \leqslant s$ and $r_{ui} = Q_{X_a}(\tau \mid X_{-a}, \varpi) - Z^a\beta_u$.

By Knight's identity (F.45), $\widehat{R}_u(\beta_u) = -\mathbb{E}_n[K_\varpi(W)r_u\int_0^1 1\{\epsilon_u \leqslant -tr_u\} - 1\{\epsilon_u \leqslant 0\}\,dt] \geqslant 0$.

$$\begin{array}{rl} \mathbb{E}_n\mathrm{E}[\widehat{R}_u(\beta_u) \mid X_{-a}, \varpi] & = \mathbb{E}_n[K_\varpi(W)r_u\int_0^1 F_{X_a|X_{-a},\varpi}(Z^a\beta_u + (1-t)r_u) - F_{X_a|X_{-a},\varpi}(Z^a\beta_u + r_u)\,dt] \\ & \leqslant \mathbb{E}_n[K_\varpi(W)r_u\int_0^1 \bar{f}tr_u dt] \leqslant \bar{f}\|r_u\|_{n,\varpi}^2/2. \leqslant C\bar{f}s/n. \end{array}$$

Since Condition CI assumes $\mathrm{E}[\|r_u\|_{n,\varpi}^2] \leqslant \mathrm{P}(\varpi)s/n$, by Markov's inequality we have $P(\widehat{R}_u(\bar{\beta}_u) \leqslant C\bar{f}s/n) \geqslant 1/2$.

Define $z_{ui} := -\int_0^1 1\{\epsilon_{ui} \leqslant -tr_{ui}\} - 1\{\epsilon_{ui} \leqslant 0\}\,dt$, so that $\widehat{R}_u(\beta_u) = \mathbb{E}_n[K_\varpi(W)r_u z_u]$ where $|z_{ui}| \leqslant 1$. We have $P(\mathbb{E}_n[K_\varpi(W)r_u z_u] \leqslant 2C\bar{f}s/n) \geqslant 1/2$ so that by Lemma 2.3.7 in [62] (note that the Lemma does not require zero mean stochastic processes), for $t \geqslant 2C\bar{f}s/n$ we have

$$\frac{1}{2}\mathrm{P}(\sup_{u\in\mathcal{U}}|\mathbb{E}_n[K_\varpi(W)r_uz_u]| \geqslant t) \leqslant 2\mathrm{P}(\sup_{u\in\mathcal{U}}|\mathbb{E}_n[K_\varpi(W)r_uz_u\epsilon]| > t/4)$$

Consider the class of functions $\mathcal{F} = \{-K_\varpi(W)r_u(1\{\epsilon_{ui} \leqslant -B_ir_{ui}\} - 1\{\epsilon_{ui} \leqslant 0\}) : u \in \mathcal{U}\}$ where $B_i \sim$ Uniform$(0,1)$ independent of $(X_i, W_i)_{i=1}^n$. It follows that $K_\varpi(W)r_uz_u = \mathrm{E}[-K_\varpi(W)r_u(1\{\epsilon_{ui} \leqslant -B_ir_{ui}\} - 1\{\epsilon_{ui} \leqslant 0\}) \mid X_i, W_i]$ where the expectation is taken over B_i only. Thus we will bound the entropy of $\overline{\mathcal{F}} = \{\mathrm{E}[f \mid X, W] : f \in \mathcal{F}\}$ via Lemma 24. Note that $\mathcal{R} := \{r_u = Q_{X_a}(\tau \mid X_{-a}, \varpi) - Z^a\beta_u : u \in \mathcal{U}\}$ where $\mathcal{G} := \{Z^a\beta_u : u \in \mathcal{U}\}$ is contained in the union of at most $|V|\binom{p}{s}$ VC-classes of dimension Cs and $\mathcal{H} := \{Q_{X_a}(\tau \mid X_{-a}, \varpi) : u \in \mathcal{U}\}\}$ is the union of $|V|$ VC-class of functions of dimension $(1+d_W)$ by Condition CI. Finally note that $\mathcal{E} := \{\epsilon_{ui} : u \in \mathcal{U}\} \subset \{X_{ai} : a \in V\} - \mathcal{G} - \mathcal{R}$.

Therefore, we have

$$\begin{aligned}
\sup\nolimits_Q \log N(\epsilon\|\bar{F}\|_{Q,2}, \overline{\mathcal{F}}, \|\cdot\|_{Q,2}) \quad &\leqslant \sup\nolimits_Q \log N((\epsilon/4)^2\|F\|_{Q,2}, \mathcal{F}, \|\cdot\|_{Q,2}) \\
&\leqslant \sup\nolimits_Q \log N(\tfrac{1}{8}(\epsilon^2/16), \mathcal{W}, \|\cdot\|_{Q,2}) \\
&+ \sup\nolimits_Q \log N(\tfrac{1}{8}(\epsilon^2/16)\|F\|_{Q,2}, \mathcal{R}, \|\cdot\|_{Q,2}) \\
&+ \sup\nolimits_Q \log N(\tfrac{1}{8}(\epsilon^2/16), 1\{\mathcal{E} + \{B\}\mathcal{R} \leqslant 0\} - 1\{\mathcal{E} \leqslant 0\}, \|\cdot\|_{Q,2})
\end{aligned}$$

We will apply Lemma 18 with envelope $\bar{F} = \sup_{u\in\mathcal{U}}|K_\varpi(W)r_u|$, so that $\mathrm{E}[\max_{i\leqslant n}\bar{F}_i^2] \leqslant C$, and $\sup_{u\in\mathcal{U}}\mathrm{E}[K_\varpi(W)r_u^2] \leqslant Cs/n =: \sigma^2$ by Condition CI. Thus, we have that with probability $1 - o(1)$

$$\sup_{u\in\mathcal{U}}|\mathbb{E}_n[K_\varpi(W)r_uz_u\epsilon]| \lesssim \sqrt{\frac{s(1+d_W)\log(p|V|n)}{n}}\sqrt{\frac{s}{n}} + \frac{s(1+d_W)\log(p|V|n)}{n} \lesssim \frac{s(1+d_W)\log(p|V|n)}{n}$$

under $M_n\sqrt{s^2/n} \leqslant C$. ∎

Lemma 3 (CIQGM, Event Ω_3)**.** *For $u = (a, \tau, \varpi) \in \mathcal{U} := V \times \mathcal{T} \times \mathcal{W}$, define the function $g_u(\delta, X, W) = K_\varpi(W)\{\rho_\tau(X_a - Z^a(\beta_u + \delta)) - \rho_\tau(X_a - Z^a\beta_u)\}$, and the event*

$$\Omega_3 := \left\{ \sup_{u\in\mathcal{U}, 1/\sqrt{n}\leqslant\|\delta\|_{1,\varpi}\leqslant\sqrt{n}} \frac{|\mathbb{E}_n[g_u(\delta, X, W) - \mathrm{E}[g_u(\delta, X, W) \mid X_{-a}, W]]|}{\|\delta\|_{1,\varpi}} < t_3 \right\}.$$

Then, under Condition CI we have $P(\Omega_3) \geqslant 1 - \gamma$ for any

$$t_3\sqrt{n} \geqslant 12 + 16\sqrt{2\log(64|V|p^2n^{3+2d_W}\log(n)L_{\mathcal{W}\mathcal{T}}^{1+d_W}\max_{a\in V, j\in[p]}\mathrm{E}[|Z_j^a|^{1+d_W/\rho}]/\gamma)}$$

Proof. We have that $\Omega_3^c := \{\max_{a\in V} A_a \geqslant t_3\sqrt{n}\}$ for

$$A_a := \sup_{(\tau,\varpi)\in\mathcal{T}\times\mathcal{W}, \underline{N}\leqslant\|\delta\|_{1,\varpi}\leqslant\bar{N}} \sqrt{n}\left|\frac{\mathbb{E}_n[g_u(\delta, X, W) - \mathrm{E}[g_u(\delta, X, W) \mid X_{-a}, W]]}{\|\delta\|_{1,\varpi}}\right|.$$

Therefore, for $\underline{N} = 1/\sqrt{n}$ and $\bar{N} = \sqrt{n}$ we have by Lemma 13

$$\begin{aligned} P(\Omega_3^c) &= P(\max_{a\in V} A_a \geqslant t_3\sqrt{n}) \\ &\leqslant |V| \max_{a\in V} P(A_a \geqslant t_3\sqrt{n}) \\ &= |V| \max_{a\in V} \mathrm{E}_{X_{-a},W} \{P(A_a \geqslant t_3\sqrt{n} \mid X_{-a}, W)\} \\ &\leqslant |V| \max_{a\in V} \mathrm{E}_{X_{-a},W} \left\{ 8p|\widehat{\mathcal{N}}| \cdot |\widehat{\mathcal{W}}| \cdot |\widehat{\mathcal{T}}| \exp(-(t_3\sqrt{n}/4-3)^2/32) \right\} \\ &\leqslant \exp(-(t_3\sqrt{n}/4-3)^2/32)|V| 64pn^{1+d_W} \log(n) L_f \mathrm{E}_{X_{-a}} \left\{ \frac{\max_{i\leqslant n} \|Z_i^a\|_\infty^{1+d_W/\rho}}{\underline{N}^{1+d_W}} \right\} \\ &\leqslant \gamma \end{aligned}$$

by the choice of t_3. ∎

Lemma 4 (CIQGM, Uniform Control of Approximation Error in Auxiliary Equation)**.** *Under Condition CI, with probability $1-o(1)$ uniformly over $u \in \mathcal{U}$ and $j \in [p]$ we have*

$$\mathbb{E}_n[K_\varpi(W) f_u^2 \{Z_{-j}^a(\gamma_u^j - \bar{\gamma}_u^j)\}^2] \quad \lesssim \underline{f}_u^2 \mathrm{P}(\varpi) \frac{s\log(p|V|n)}{n}$$

Proof. Define the class of functions $\mathcal{G} = \cup_{a\in V, j\in[p]} \mathcal{G}_{aj}$ with $\mathcal{G}_{aj} := \{Z_{-j}^a(\gamma_u^j - \bar{\gamma}_u^j) : \tau \in \mathcal{T}, \varpi \in \mathcal{W}\}$. Under Condition CI we have $\sup_{u\in\mathcal{U}} \|\bar{\gamma}_u^j\|_0 \leqslant Cs$, $\sup_{u\in\mathcal{U}, j\in[p]} \|\bar{\gamma}_u^j - \gamma_u^j\| \vee \frac{\|\bar{\gamma}_u^j - \gamma_u^j\|_1}{\sqrt{s}} \leqslant \{n^{-1} s \log(p|V|n)\}^{1/2}$. Without loss of generality we can take $\|\bar{\gamma}_u^j - \bar{\gamma}_{u'}^j\| \leqslant \|\gamma_u^j - \gamma_{u'}^j\|$. By Lemma 5, we have $\|\gamma_u^j - \gamma_{u'}^j\| \leqslant L_\gamma(\|u-u'\| + \|u-u'\|^{1/2})$ for each $a \in V$, $j \in [p]$. Therefore,

$$\begin{aligned} \|\{Z_{-j}^a(\bar{\gamma}_u^j - \gamma_u^j)\}^2 - \{Z_{-j}^a(\bar{\gamma}_{u'}^j - \gamma_{u'}^j)\}^2\|_{Q,2} &\leqslant \|Z_{-j}^a(\bar{\gamma}_u^j - \bar{\gamma}_{u'}^j + \gamma_{u'}^j - \gamma_u^j) Z_{-j}^a(\bar{\gamma}_u^j - \gamma_u^j + \bar{\gamma}_{u'}^j - \gamma_{u'}^j)\|_{Q,2} \\ &\leqslant \|\|Z_{-j}^a\|_\infty^2\|_{Q,2} \|\bar{\gamma}_u^j - \bar{\gamma}_{u'}^j + \gamma_{u'}^j - \gamma_u^j\|_1 \|\bar{\gamma}_u^j - \gamma_u^j + \bar{\gamma}_{u'}^j - \gamma_{u'}^j\|_1 \\ &\leqslant 4\|\|Z_{-j}^a\|_\infty^2\|_{Q,2} \sup_{u\in\mathcal{U}} \|\bar{\gamma}_u^j - \gamma_u^j\|_1 \sqrt{2p} \|\gamma_u^j - \gamma_{u'}^j\| \\ &\leqslant \|\|Z_{-j}^a\|_\infty^2\|_{Q,2} L'_\gamma(\|u-u'\| + \|u-u'\|^{1/2}). \end{aligned}$$

where $L'_\gamma = 4\{n^{-1}s^2\log(p|V|n)\}^{1/2}\sqrt{2p}L_\gamma$. Thus, for the envelope $G = \max_{a\in V} \|Z^a\|_\infty^2 \sup_{u\in\mathcal{U}} \|\bar{\gamma}_u^j - \gamma_u^j\|_1^2$ that

$$\log N(\epsilon\|G\|_{Q,2}, \mathcal{G}, \|\cdot\|_{Q,2}) \leqslant \log(|V|p) + \log N(\epsilon \tfrac{\sup_{u\in\mathcal{U}} \|\bar{\gamma}_u^j - \gamma_u^j\|_1^2}{L'_\gamma}, \mathcal{U}, d_{\mathcal{U}}) \leqslant C(1+d_W)^2 \log(L'_\gamma n/\epsilon).$$

Next define the functions $\mathcal{W}_0 = \{K_\varpi(W) f_u^2 : u \in \mathcal{U}\}$, $\mathcal{W}_1 = \{\mathrm{P}(\varpi)^{-1} : \varpi \in \mathcal{W}\}$ and $\mathcal{W}_2 = \{K_\varpi(W) : \varpi \in \mathcal{W}\}$. We have that $\mathcal{W}_2$ is VC class with VC index Cd_W and $\mathcal{W}_1$ is bounded by $\mu_{\mathcal{W}}^{-1}$ and covering number bounded by $(Cd_W/\{\mu_{\mathcal{W}}\epsilon\})^{1+d_W}$. Finally, since $|K_\varpi(W)f_u^2 - K_{\varpi'}(W)f_{u'}^2| \leqslant K_\varpi(W)K_{\varpi'}(W)|f_u^2 - f_{u'}^2| + \bar{f}^2|K_\varpi(W) - K_{\varpi'}(W)| \leqslant 2\bar{f}L_f\|u-u'\| + \bar{f}^2|K_\varpi(W) - K_{\varpi'}(W)|$, we have $N(\epsilon, \mathcal{U}, |\cdot|) \leqslant (C(1+d_W)/\epsilon)^{1+d_W}$. Therefore, using standard bounds we have

$$\log N(\epsilon\|\mu_{\mathcal{W}}^{-1} G\bar{f}\|_{Q,2}, \mathcal{W}_0\mathcal{W}_1\mathcal{W}_2\mathcal{G}, \|\cdot\|_{Q,2}) \lesssim (1+d_W)^2 \log(L'_\gamma L_f n/\epsilon)$$

By Theorem 5.1 in [25] we have

$$\begin{aligned} &\sup_{u\in\mathcal{U}, j\in[p]} |(\mathbb{E}_n - \mathrm{E})[f_u^2\{Z_{-j}^a(\gamma_u^j - \bar{\gamma}_u^j)\}^2/\mathrm{P}(\varpi)]| \\ &\lesssim \sqrt{\frac{(1+d_W)^2\log(p|V|n)\sup_{u\in\mathcal{U}} \mathrm{E}[K_\varpi(W)f_u^4\{Z_{-j}^a(\gamma_u^j - \bar{\gamma}_u^j)\}^4]/\mathrm{P}(\varpi)^2}{n}} + \frac{(1+d_W)^2 M_n^2 \mu_{\mathcal{W}}^{-1} \sup_{u\in\mathcal{U}} \|\bar{\gamma}_u^j - \gamma_u^j\|_1^2 \log(p|V|n)}{n} \\ &\lesssim \sqrt{\frac{(1+d_W)^2\log(p|V|n)}{\mu_{\mathcal{W}} n} \frac{s\log(p|V|n)}{n}} + \frac{(1+d_W)^2 \delta_n^2 \underline{f}_{\mathcal{U}}^2 \mu_{\mathcal{W}} s \log(p|V|n)}{n} \end{aligned}$$

where we used that $\mathrm{E}[f_u^4\{Z^a\xi\}^4 \mid \varpi] \leqslant \bar{f}^4\mathrm{E}[\{Z^a\xi\}^4 \mid \varpi] \leqslant C\|\xi\|^4$, $\|\bar{\gamma}_u^j - \gamma_u^j\| + s^{-1/2}\|\bar{\gamma}_u^j - \gamma_u^j\|_1 \leqslant \{n^{-1}s\log(p|V|n)\}^{1/2}$, $(1+d_W)\log(p|V|n) \leqslant \delta_n n \underline{f}_{\mathcal{U}}^4 \mu_{\mathcal{W}}^3$ and $M_n s \log^{1/2}(p|V|n) \leqslant \delta_n n^{1/2}\mu_{\mathcal{W}}\underline{f}_{\mathcal{U}}$ by Condition CI. Furthermore, by Condition CI, the result follows from $\mathrm{E}[f_u^2\{Z_{-j}^a(\bar{\gamma}_u^j - \gamma_u^j)\}^2 \mid \varpi] \leqslant C\underline{f}_u^2\|\bar{\gamma}_u^j - \gamma_u^j\|^2 \leqslant C\underline{f}_u^2 n^{-1}s\log(p|V|n)$. ∎

Lemma 5. *Under Condition CI, for $u = (a,\tau,\varpi) \in \mathcal{U}$ and $u' = (a,\tau',\varpi') \in \mathcal{U}$ we have that*

$$\|\gamma_u^j - \gamma_{u'}^j\| \leqslant \frac{C'}{\underline{f}_{u'}^2 \mathrm{P}(\varpi')}\{\mathrm{E}[\{K_{\varpi'}(W) - K_\varpi(W)\}^2]^{1/2} + \mathrm{E}[K_\varpi(W)K_{\varpi'}(W)\{f_{u'}^2 - f_u^2\}^2]^{1/2}\}.$$

In particular, we have $\|\gamma_u^j - \gamma_{u'}^j\| \leqslant L_\gamma\{\|\varpi - \varpi'\|^{1/2} + \|u - u'\|\}$ for $L_\gamma = C'\{L' + L\}/\{\underline{f}_{\mathcal{U}}^2\mu_{\mathcal{W}}\}$ under $\mathrm{E}[|K_\varpi(W) - K_{\varpi'}(W)|] \leqslant L\|\varpi - \varpi'\|$, $K_\varpi(W)K_{\varpi'}(W)|f_{u'} - f_u| \leqslant L'\|u' - u\|$, and $f_u \leqslant \bar{f}$.

Proof. Let $u = (a,\tau,\varpi)$ and $u' = (a,\tau',\varpi')$. By Condition CI we have

$$\|\gamma_u^j - \gamma_{u'}^j\|^2 \quad \leqslant C\mathrm{E}[\{Z_{-j}^a(\gamma_u^j - \gamma_{u'}^j)\}^2 \mid \varpi] \leqslant \{C/\mathrm{P}(\varpi)\}\mathrm{E}[K_{\varpi'}(W)\{Z_{-j}^a(\gamma_u^j - \gamma_{u'}^j)\}^2]$$

To bound the last term of the right hand side above, by definition of $\underline{f}_{u'}$ and using Cauchy-Schwarz's inequality we have

$$\begin{aligned}\underline{f}_{u'}\mathrm{E}[K_{\varpi'}(W)\{Z_{-j}^a(\gamma_u^j - \gamma_{u'}^j)\}^2] &\leqslant \mathrm{E}[K_{\varpi'}(W)f_{u'}\{Z_{-j}^a(\gamma_u^j - \gamma_{u'}^j)\}^2] \\ &\leqslant \{\mathrm{E}[K_{\varpi'}(W)f_{u'}^2\{Z_{-j}^a(\gamma_u^j - \gamma_{u'}^j)\}^2]\ \mathrm{E}[K_{\varpi'}(W)\{Z_{-j}^a(\gamma_u^j - \gamma_{u'}^j)\}^2]\}^{1/2}\end{aligned}$$

so that $\mathrm{E}[K_{\varpi'}(W)\{Z_{-j}^a(\gamma_u^j - \gamma_{u'}^j)\}^2]^{1/2} \leqslant \{\mathrm{E}[K_{\varpi'}(W)f_{u'}^2\{Z_{-j}^a(\gamma_u^j - \gamma_{u'}^j)\}^2]\}^{1/2}/\underline{f}_{u'}$. Therefore

$$\|\gamma_u^j - \gamma_{u'}^j\|^2 \leqslant \{1/\underline{f}_{u'}\}^2\{C/\mathrm{P}(\varpi)\}\mathrm{E}[K_{\varpi'}(W)f_{u'}^2\{Z_{-j}^a(\gamma_u^j - \gamma_{u'}^j)\}^2]. \tag{E.39}$$

We proceed to bound the last term. The optimality of γ_u^j and $\gamma_{u'}^j$ yields

$$\mathrm{E}[K_\varpi(W)f_u^2 Z_{-j}^a(Z_j^a - Z_{-j}^a\gamma_u^j)] = 0 \quad \text{and} \quad \mathrm{E}[K_{\varpi'}(W)f_{u'}^2 Z_{-j}^a(Z_j^a - Z_{-j}^a\gamma_{u'}^j)] = 0$$

Therefore, we have

$$\begin{aligned}\mathrm{E}[K_{\varpi'}(W)f_{u'}^2\{Z_{-j}^a(\gamma_u^j - \gamma_{u'}^j)\}Z_{-j}^a] &= -\mathrm{E}[K_{\varpi'}(W)f_{u'}^2\{Z_j^a - Z_{-j}^a\gamma_u^j\}Z_{-j}^a] \\ &= -\mathrm{E}[\{K_{\varpi'}(W)f_{u'}^2 - K_\varpi(W)f_u^2\}\{Z_j^a - Z_{-j}^a\gamma_u^j\}Z_{-j}^a]\end{aligned} \tag{E.40}$$

Multiplying by $(\gamma_u^j - \gamma_{u'}^j)$ both sides of (E.40), we have

$$\begin{aligned}&\mathrm{E}[K_{\varpi'}(W)f_{u'}^2\{Z_{-j}^a(\gamma_u^j - \gamma_{u'}^j)\}^2] \\ &\leqslant \mathrm{E}[\{K_{\varpi'}(W)f_{u'}^2 - K_\varpi(W)f_u^2\}^2]^{1/2}\{\mathrm{E}[\{Z_j^a - Z_{-j}^a\gamma_u^j\}^2\{Z_{-j}^a(\gamma_u^j - \gamma_{u'}^j)\}^2]\}^{1/2} \\ &\leqslant \mathrm{E}[\{K_{\varpi'}(W)f_{u'}^2 - K_\varpi(W)f_u^2\}^2]^{1/2}C\|\gamma_u^j - \gamma_{u'}^j\|\end{aligned}$$

by the fourth moment assumption in Condition CI. By Condition CI, $f_u, f_{u'} \leqslant \bar{f}$, and it follows that

$$|K_\varpi(W)f_u^2 - K_{\varpi'}(W)f_{u'}^2| \leqslant K_\varpi(W)K_{\varpi'}(W)|f_u^2 - f_{u'}^2| + \bar{f}^2|K_\varpi(W) - K_{\varpi'}(W)| \tag{E.41}$$

From (E.39) we obtain

$$\|\gamma_u^j - \gamma_{u'}^j\| \leqslant \frac{C'}{\underline{f}_{u'}^2 \mathrm{P}(\varpi')}\{\mathrm{E}[\{K_{\varpi'}(W) - K_\varpi(W)\}^2]^{1/2} + \mathrm{E}[K_\varpi(W)K_{\varpi'}(W)\{f_{u'}^2 - f_u^2\}^2]^{1/2}\}.$$

∎

Lemma 6. *Let $\mathcal{U} = V \times \mathcal{T} \times \mathcal{W}$. Under Condition CI, and $K_\varpi(W)K_{\varpi'}|f_u - f_{u'}| \leqslant L_f\|u-u'\|$, for $m = 1,2$, we have*

$$\mathrm{E}\left[\sup_{u\in\mathcal{U},\|\theta\|_0\leqslant k,\|\theta\|=1}|(\mathbb{E}_n-\mathrm{E})[K_\varpi(W)f_u^m(Z^a\theta)^2]|\right] \lesssim C\delta_n \sup_{u\in\mathcal{U},\|\theta\|_0\leqslant k,\|\theta\|=1}\{\mathrm{E}[K_\varpi(W)f_u^m(Z^a\theta)^2]\}^{1/2}$$

where $\delta_n = M_n\sqrt{k}\sqrt{\{1+d_W\}C\log(np|V|)}\log(1+k)\sqrt{\log n}/\sqrt{n}$. Moreover, under Condition CI, $\delta_n = o(\mu_{\mathcal{W}})$.

Proof. By symmetrization we have

$$\mathrm{E}[\sup_{u\in\mathcal{U},\|\theta\|_0\leqslant k,\|\theta\|=1}|(\mathbb{E}_n-\mathrm{E})[K_\varpi(W)f_u^m(Z^a\theta)^2]|] \leqslant 2\mathrm{E}[\sup_{u\in\mathcal{U},\|\theta\|_0\leqslant k,\|\theta\|=1}|\mathbb{E}_n[\epsilon K_\varpi(W)f_u^m(Z^a\theta)^2]|]$$

where ϵ_i are i.i.d. Radamacher random variables. Further, conditional on $\{(W_i,X_i), i=1,\ldots,n\}$, $\{K_\varpi(W_i) : i=1,\ldots,n, \varpi\in\mathcal{W}\}$ induces at most n^{d_W} different sequences by Corollary 2.6.3 in [61]. Moreover, $|K_\varpi f_u^m - K_{\varpi'}f_{u'}^m| \leqslant K_\varpi K_{\varpi'}|f_u-f_{u'}|\{1+2\bar{f}\}+\bar{f}^m|K_\varpi - K_{\varpi'}|$ for $m=1,2$ where u and u' have the same $a\in V$. We can take a cover $\widehat{\mathcal{U}}$ of $V\times\mathcal{T}\times\mathcal{W}$ such that $\|u-u'\| \leqslant \{L_f(1+2\bar{f})n\max_{i\leqslant n}\|Z_i^a\|_\infty^2 k\}^{-1}$ so that

$$\begin{aligned}\Big|\sup\nolimits_{u\in\mathcal{U},\|\theta\|_0\leqslant k,\|\theta\|=1}|\mathbb{E}_n[\epsilon K_\varpi(W)f_u^m(Z^a\theta)^2]| - \sup\nolimits_{u\in\widehat{\mathcal{U}},\|\theta\|_0\leqslant k,\|\theta\|=1}|\mathbb{E}_n[\epsilon K_\varpi(W)f_u^m(Z^a\theta)^2]|\Big| \\ \leqslant \mathbb{E}_n[K_\varpi(W)K_{\varpi'}(W)n^{-1}]\end{aligned}$$

so that

$$\mathrm{E}\left[\sup_{u\in\mathcal{U},\|\theta\|_0\leqslant k,\|\theta\|=1}|\mathbb{E}_n[\epsilon K_\varpi(W)f_u^m(Z^a\theta)^2]|\right] \leqslant \mathrm{E}_{W,X}\mathrm{E}_\epsilon\left[\sup_{u\in\widehat{\mathcal{U}},\|\theta\|_0\leqslant k,\|\theta\|=1}|\mathbb{E}_n[\epsilon K_\varpi(W)f_u^m(Z^a\theta)^2]|\right] + \frac{1}{n}$$

and $|\widehat{\mathcal{U}}| \leqslant |V|n^{d_W}\{L_f(1+2\bar{f})nk\max_{i\leqslant n}\|Z_i^a\|_\infty^2\}^{(1+d_W)}$.

By Lemma 17 with $K = K(W,X) = (1+\bar{f}^2)\sup_{a\in V}\max_{i\leqslant n}\|Z^a\|_\infty$ and

$$\begin{aligned}\delta_n(W,X) \quad &:= \bar{C}K(W,X)\sqrt{k}\left(\sqrt{\log|\widehat{\mathcal{U}}|}+\sqrt{1+\log p}+\log k\sqrt{\log(p\vee n)}\sqrt{\log n}\right)/\sqrt{n} \\ &\lesssim K(W,X)\sqrt{k}\sqrt{\{1+d_W\}C\log(np|V|K(W,X))}\log(1+k)\sqrt{\log n}/\sqrt{n}\end{aligned}$$

so that conditional on (W,X) we have

$$\mathrm{E}_\epsilon\left[\sup_{u\in\widehat{\mathcal{U}},\|\theta\|_0\leqslant k,\|\theta\|=1}|\mathbb{E}_n[\epsilon K_\varpi(W)f_u^m(Z^a\theta)^2]|\right] \lesssim \delta_n(W,X)\sup_{u\in\widehat{\mathcal{U}},\|\theta\|_0\leqslant k,\|\theta\|=1}\sqrt{\mathbb{E}_n[K_\varpi(W)f_u^m(Z^a\theta)^2]}$$

Therefore,

$$\begin{aligned}&\mathrm{E}\left[\sup_{u\in\mathcal{U},\|\theta\|_0\leqslant k,\|\theta\|=1}|\mathbb{E}_n[\epsilon K_\varpi(W)f_u^m(Z^a\theta)^2]|\right] \\ &\leqslant \mathrm{E}_{W,X}\left[\delta_n(W,X)\sup_{u\in\widehat{\mathcal{U}},\|\theta\|_0\leqslant k,\|\theta\|=1}\sqrt{\mathbb{E}_n[K_\varpi(W)f_u^m(Z^a\theta)^2]}\right] \\ &\leqslant \mathrm{E}_{W,X}[\delta_n^2(W,X)] + \mathrm{E}_{W,X}[\delta_n^2(W,X)]^{1/2}\sup_{u\in\widehat{\mathcal{U}},\|\theta\|_0\leqslant k,\|\theta\|=1}\mathrm{E}[K_\varpi(W)f_u^m(Z^a\theta)^2]^{1/2}\end{aligned}$$

Note that for a random variable $A\geqslant 1$, we have that $\mathrm{E}[A^2\sqrt{\log(CA)}] \leqslant \mathrm{E}[A^2]\sqrt{\log(C)}+\mathrm{E}[A^2\sqrt{\log(A)}] \leqslant \mathrm{E}[A^2]\sqrt{\log(C)}+\mathrm{E}[A^{2+1/4}]$. Therefore, under Condition CI, since $q\geqslant 2+1/4$, we have

$$\mathrm{E}_{W,X}[\delta_n^2(W,X)]^{1/2} \lesssim M_n\sqrt{k}\sqrt{\{1+d_W\}C\log(np|V|)}\log(1+k)\sqrt{\log n}/\sqrt{n}$$

■

E.1. **Results for Predictive Quantile Graph Models.** In the analysis of the PQGM we also use the following event for some sequence $(K_u)_{u\in\mathcal{U}}$

$$\begin{aligned}\Omega_4 = \{K_u\{\mathbb{E}_n[K_w(W)X_j^2]\}^{1/2} \geqslant \mathbb{E}_n[\mathrm{E}[K_\varpi(W)(\tau - 1\{X_a \leqslant X_{-a}\beta_u + r_u\})X_{-a} \mid X_j, W]],\\ \text{for all } u \in \mathcal{U}, j \in V \setminus \{a\}\}.\end{aligned} \tag{E.42}$$

Lemma 7 (Rate for PQGM)**.** *Suppose that Ω_1, Ω_2, Ω_3 and Ω_4 hold. Further assume $2\frac{1+1/c}{1-1/c}\|\beta_u\|_{1,\varpi} + \frac{1}{\lambda_u(1-1/c)}\bar{R}_{u\gamma} \leqslant \sqrt{n}$ for all $u \in \mathcal{U}$, and (F.44) holds for all $\delta \in A_u := \Delta_{\varpi,2\mathbf{c}} \cup \{v : \|v\|_{1,\varpi} \leqslant 2\mathbf{c}\bar{R}_{u\gamma}/\lambda_u\}$, $q_{A_u}/4 \geqslant (\sqrt{\bar{f}}+1)\|r_u\|_{n,\varpi} + [\lambda_u + t_3 + K_u]\frac{3\mathbf{c}\sqrt{s}}{\kappa_{u,2\mathbf{c}}}$ and $q_{A_u} \geqslant \{2\mathbf{c}\left(1+\frac{t_3+K_u}{\lambda_u}\right)\bar{R}_{u\gamma}\}^{1/2}$. Then uniformly over all $u = (a,\tau,\varpi) \in \mathcal{U} := V \times \mathcal{T} \times \mathcal{W}$ the $\|\cdot\|_{1,\varpi}$-penalized estimator $\widehat{\beta}_u$ satisfies*

$$\begin{aligned}\|\sqrt{f_u}X_{-a}(\widehat{\beta}_u - \beta_u)\|_{n,\varpi} &\leqslant \sqrt{8\mathbf{c}\left(1+\frac{t_3}{\lambda_u}\right)\bar{R}_{u\gamma}} + (\bar{f}^{1/2}+1)\|r_u\|_{n,\varpi} + [\lambda_u + t_3 + K_u]\tfrac{3\mathbf{c}\sqrt{s}}{\kappa_{u,2\mathbf{c}}}\\ \|\widehat{\beta}_u - \beta_u\|_{1,\varpi} &\leqslant (1+2\mathbf{c})\sqrt{s}\|\sqrt{f_u}X_{-a}\delta_u\|_{n,\varpi}/\kappa_{u,2\mathbf{c}} + \tfrac{2\mathbf{c}}{\lambda_u}\bar{R}_{u\gamma}\end{aligned}$$

Proof of Lemma 7. The proof proceeds similarly to the proof of Lemma 1 by defining

$$\widehat{R}_u(\beta) = \mathbb{E}_n[K_\varpi(W)\{\rho_u(X_a - X_{-a}\beta) - \rho_u(X_a - X_{-a}\beta_u - r_u) - (\tau - 1\{X_a \leqslant X_{-a}\beta_u + r_u\})(X_{-a}\beta - X_{-a}\beta_u - r_u)\}].$$

The same argument yields $\delta_u = \widehat{\beta}_u - \beta_u \in A_u := \Delta_{\varpi,2\mathbf{c}} \cup \{v : \|v\|_{1,\varpi} \leqslant 2\mathbf{c}\bar{R}_{u\gamma}/\lambda_u\}$ under $\Omega_1 \cap \Omega_2$. (Similarly we also have $\|\delta_u\|_{1,\varpi} \leqslant \sqrt{n}$.) Furthermore, under $\Omega_1 \cap \Omega_2 \cap \Omega_3$ we also have (E.38), in particular

$$\mathbb{E}_n\mathrm{E}[K_\varpi(W)\{\rho_u(X_a - X_{-a}(\beta_u + \delta_u)) - \rho_u(X_a - X_{-a}\beta_u)\} \mid X_{-a}, W] \leqslant \lambda_u\|\delta_u\|_{1,\varpi} + t_3\|\delta_u\|_{1,\varpi}$$

Since the conditions of Lemma 12 hold we have

$$\begin{aligned}\mathbb{E}_n\mathrm{E}[K_\varpi(W)\{\rho_\tau(X_a - X_{-a}(\beta_u + \delta_u)) - \rho_\tau(X_a - X_{-a}\beta_u)\} \mid X_{-a}, W]\\ \geqslant -(\sqrt{\bar{f}}+1)\|r_u\|_{n,\varpi}\|\sqrt{f_u}X_{-a}\delta\|_{n,\varpi} - K_u\|\delta_u\|_{1,\varpi}\\ + \frac{\|\sqrt{f_u}X_{-a}\delta_u\|_{n,\varpi}^2}{4} \wedge \bar{q}_{A_u}\|\sqrt{f_u}X_{-a}\delta_u\|_{n,\varpi}\end{aligned}$$

where K_u is given in Ω_4 which accounts for the misspecification the conditional quantile condition. Therefore, we have

$$\begin{aligned}\frac{\|\sqrt{f_u}X_{-a}\delta_u\|_{n,\varpi}^2}{4} \wedge \bar{q}_{A_u}\|\sqrt{f_u}X_{-a}\delta_u\|_{n,\varpi} &\leqslant (\bar{f}^{1/2}+1)\|r_u\|_{n,\varpi}\|\sqrt{f_u}X_{-a}\delta\|_{n,\varpi} + \{\lambda_u + t_3 + K_u\}\|\delta_u\|_{1,\varpi}\\ &\leqslant \{(\bar{f}^{1/2}+1)\|r_u\|_{n,\varpi} + \tfrac{3\mathbf{c}\sqrt{s}}{\kappa_{u,2\mathbf{c}}}(K_u + t_3 + \lambda_u)\}\|\sqrt{f_u}X_{-a}\delta_u\|_{n,\varpi}\\ &\quad + \{\lambda_u + t_3 + K_u\}\tfrac{2\mathbf{c}}{\lambda_u}\bar{R}_{u\gamma}\end{aligned}$$

The result then follows with the same argument under the current assumptions that account for K_u. ■

Lemma 8 (PQGM, Event Ω_1)**.** *Under Condition P, we have*

$$\mathrm{P}\left(\sup_{u\in\mathcal{U},j\in V}\frac{|\mathbb{E}_n[K_\varpi(W)(\tau - 1\{X_a \leqslant X_{-a}'\beta_u + r_u\})X_j]|}{\{\mathbb{E}_n[K_\varpi(W)X_j^2]\}^{1/2}} > t\right) \leqslant 2\{ne/d_W\}^{2d_W}\exp(-\{t/[4(1+\delta_n)]\}^2)$$

where $t \geqslant 4\sup_{u\in\mathcal{U}}\{\bar{\mathrm{E}}[K_\varpi(W)(\tau - 1\{X_a \leqslant X_{-a}'\beta_u\})^2X_j^2]\}^{1/2}$ and $\delta_n = o(1)$. In particular, the RHS is less than ξ if $t \geqslant 2(1+1/16)\sqrt{2\log(8|V|^2\{ne/d_W\}^{2d_W}/\xi)}$.

Proof. We have that

$$\begin{array}{rl} & \mathrm{P}(\lambda_0 \leqslant \sup_{u\in\mathcal{U},j\in V\setminus\{a\}} |\mathbb{E}_n[K_\varpi(W)(\tau - 1\{X_a \leqslant X_{-a}'\beta_u\})X_j]|/\mathbb{E}_n[K_\varpi(W_i)X_j^2]^{1/2}) \\ \leqslant & \mathrm{P}(\lambda_0 \leqslant (1+\delta) \sup_{u\in\mathcal{U},j\in V\setminus\{a\}} |\mathbb{E}_n[K_\varpi(W)(\tau - 1\{X_a \leqslant X_{-a}'\beta_u\})X_j]|/\mathrm{E}[K_\varpi(W)X_j^2]^{1/2}) \\ & +\mathrm{P}(\sup_{u\in\mathcal{U},j\in V} \mathrm{E}[K_\varpi(W)X_j^2]^{1/2}/\mathbb{E}_n[K_\varpi(W)X_j^2]^{1/2} \geqslant (1+\delta)) \end{array}$$

Under Condition P, $c\mu_{\mathcal{W}} \leqslant \mathrm{E}[K_\varpi(W)X_j^2] \leqslant C$ and $\mathcal{W}$ is a VC class of set with VC dimension d_W. Therefore, by Lemma 18 we have that with probability $1-o(1)$

$$\sup_{u\in\mathcal{U},j\in V} (\mathbb{E}_n - \mathrm{E})[K_\varpi(W)X_j^2] \lesssim \sqrt{\frac{(1+d_W)\log(|V|M_n/\sigma)}{n}} + \frac{(1+d_W)M_n^2\log(|V|M_n/\sigma)}{n}$$

for $\sigma^2 = \max_{j\in V,\varpi\in\mathcal{W}} \mathrm{E}[K_\varpi(W)X_j^2] \leqslant \max_{j\in V}\mathrm{E}[X_j^2] \leqslant C$ and envelope $F = \|X\|_\infty^2$ so that $\|F\|_{P,2} \leqslant \|\max_{i\leqslant n} F_i\|_{P,2} \leqslant M_n^2$. Thus for $\delta_n \to 0$, provided $M_n^2\log(n|V|) = o(n^{1/2})$ and $(1+d_W) = o(n\delta_n^2\mu_{\mathcal{W}}^2)$,

$$\mathrm{P}\left(\frac{1}{1+\delta_n} \leqslant \frac{\mathrm{E}[K_\varpi(W)X_j^2]^{1/2}}{\mathbb{E}_n[K_\varpi(W)X_j^2]^{1/2}} \leqslant (1+\delta_n), \text{ for all } u\in\mathcal{U}, j\in V\right) = 1-o(1). \tag{E.43}$$

Set $\sigma_{uj} := \mathrm{E}[K_\varpi(W)X_j^2]^{1/2}$ and let $\sigma^2 = \sup_{u\in\mathcal{U},j\in V}\mathrm{var}(K_\varpi(W)(\tau - 1\{X_a \leqslant X_{-a}'\beta_u\})X_j/\sigma_{uj}) \leqslant 1$. By symmetrization (adapting Lemma 2.3.7 to replace the "arbitrary" factor 2 with $1+\delta_n$), for $\delta := 1/2\{1+n\sigma^2/t\} < 1/2$ we have

$$\begin{array}{l} (*) := \mathrm{P}(\sup_{u\in\mathcal{U},j\in V\setminus\{a\}} |\sum_{i=1}^n K_\varpi(W_i)(\tau - 1\{X_{ia} \leqslant X_{i,-a}'\beta_u\})X_{ij}/\sigma_{uj}| \geqslant t) \\ \leqslant 2\mathrm{P}(\sup_{u\in\mathcal{U},j\in V\setminus\{a\}} |\sum_{i=1}^n \epsilon_i K_\varpi(W_i)(\tau - 1\{X_{ia} \leqslant X_{i,-a}'\beta_u\})X_{ij}/\sigma_{uj}| \geqslant t\delta) \\ \leqslant 2\mathrm{P}\left(\sup_{u\in\mathcal{U},j\in V\setminus\{a\}} \frac{|\sum_{i=1}^n \epsilon_i K_\varpi(W_i)(\tau - 1\{X_{ia}\leqslant X_{i,-a}'\beta_u\})X_{ij}|}{\mathbb{E}_n[K_\varpi(W)X_j^2]^{1/2}} \geqslant t\delta/(1+\delta_n)\right) + o(1) \end{array}$$

where the last inequality follows from (E.43).

Therefore, by the union bound and symmetry, and iterated expectations we have

$$(*) \leqslant 4|V|\max_{j\in V]} \mathrm{E}_{W,X}[\mathrm{P}\left(\sup_{u\in\mathcal{U}} \frac{\sum_{i=1}^n \epsilon_i K_\varpi(W_i)(\tau - 1\{X_{ia} \leqslant X_{i,-a}'\beta_u\})X_{ij}}{\mathbb{E}_n[K_\varpi(W)X_j^2]^{1/2}} \geqslant t\delta/(1+\delta_n) \mid W, X\right)]$$

Next we use that $\{\varpi \in \mathcal{W}\}$ is a VC class of sets with VC dimension bounded by d_W and $\{1\{X_a \leqslant X_{-a}'\beta_u\} : (\tau,\varpi) \in \mathcal{T}\times\mathcal{W}\}$ is a VC class of sets with VC dimension bounded by $1+d_W$. By Corollary 2.6.3 in [61], we have that conditionally on $(W_i, X_i)_{i=1}^n$, the set of (binary) sequences $\{(K_\varpi(W_i))_{i=1,\ldots,n} : \varpi \in \mathcal{W}\}$ has at most $\sum_{j=0}^{d_W-1}\binom{n}{j}$ different values. Similarly, $\{(1\{X_{ia} \leqslant X_{i,-a}'\beta_u\})_{i=1,\ldots,n} : u \in \mathcal{U}\}$ assumes at most $\sum_{j=0}^{d_W}\binom{n}{j}$ different values. Assuming that $n \geqslant d_W$, we have $\sum_{j=0}^{d_W-1}\binom{n}{j} \leqslant \{ne/(d_W-1)\}^{d_W-1}$ and

$$\begin{array}{l} \mathrm{P}_\varepsilon\left(\sup_{u\in\mathcal{U}} \frac{\sum_{i=1}^n \epsilon_i K_\varpi(W_i)(\tau - 1\{X_{ia}\leqslant X_{i,-a}'\beta_u\})X_{ij}}{\mathbb{E}_n[K_\varpi(W)X_j^2]^{1/2}} \geqslant t\delta/(1+\delta_n) \mid W, X\right) \\ \leqslant |V|\{\frac{ne}{d_W-1}\}^{d_W-1}\{\frac{ne}{d_W}\}^{d_W} \sup_{u\in\mathcal{U}} \mathrm{P}_\varepsilon\left(\sup_{\tilde\tau\in\mathcal{T}} \frac{\sum_{i=1}^n \epsilon_i K_\varpi(W_i)(\tilde\tau - 1\{X_{ia}\leqslant X_{i,-a}'\beta_u\})X_{ij}}{\mathbb{E}_n[K_\varpi(W)X_j^2]^{1/2}} \geqslant t\delta/(1+\delta_n) \mid W, X\right) \\ \leqslant 2|V|\{ne/d_W\}^{2d_W} \sup_{u\in\mathcal{U},\tilde\tau\in\{\underline{\tau},\bar\tau\}} \mathrm{P}_\varepsilon\left(\frac{\sum_{i=1}^n \epsilon_i K_\varpi(W_i)(\tilde\tau - 1\{X_{ia}\leqslant X_{i,-a}'\beta_u\})X_{ij}}{\mathbb{E}_n[K_\varpi(W)X_j^2]^{1/2}} \geqslant t\delta/(1+\delta_n) \mid W, X\right) \\ \leqslant 2|V|\{ne/d_W\}^{2d_W}\exp(-\{t\delta/[1+\delta_n]\}^2) \end{array}$$

where we used that the expression is linear in τ and so it is maximized at the extremes. Therefore, setting $\lambda_0 = ct/n$ where $t \geqslant 4\sqrt{n}\sigma$ and $t \geqslant 2(1+1/16)\sqrt{2\log(8p|V|\{ne/d_W\}^{2d_W}/\xi)}$. ■

Lemma 9 (PQGM, Event Ω_2)**.** *Under Condition P we have* $\mathbb{E}_n[\mathrm{E}[\widehat{R}_u(\bar\beta_u)]] \leqslant \bar f \|r_{ui}\|_{n,\varpi}^2/2$*,* $\widehat{R}_u(\bar\beta_u) \geqslant 0$ *and*

$$\mathrm{P}(\sup_{u\in\mathcal{U}} \widehat{R}_u(\bar\beta_u) \geqslant C\{1+\bar f\}\{n^{-1}s(1+d_W)\log(p|V|n)\}) = 1-o(1).$$

Proof of Lemma 9. We have that $\widehat{R}_u(\bar\beta_u) \geqslant 0$ by convexity of ρ_τ. Let $\epsilon_{ui} = X_{ai} - X_{-ai}\bar\beta_u - r_{ui}$ where $\|\bar\beta_u\|_0 \leqslant s$ and $r_{ui} = X_{-a}(\beta_u - \bar\beta_u)$. By Knight's identity (F.45), $\widehat{R}_u(\bar\beta_u) = -\mathbb{E}_n[K_\varpi(W) r_u \int_0^1 1\{\epsilon_u \leqslant -tr_u\} - 1\{\epsilon_u \leqslant 0\}\, dt] \geqslant 0$.

$$\begin{aligned}\mathbb{E}_n\mathrm{E}[\widehat{R}_u(\bar\beta_u)] &= \mathbb{E}_n[K_\varpi(W) r_u \textstyle\int_0^1 F_{X_a|X_{-a},\varpi}(X_{-a}\bar\beta_u + (1-t)r_u) - F_{X_a|X_{-a},\varpi}(X_{-a}\bar\beta_u + r_u)\, dt]\\ &\leqslant \mathbb{E}_n\mathrm{E}[K_\varpi(W) r_u \textstyle\int_0^1 \bar f t r_u dt] \leqslant \bar f \mathrm{E}[\|r_u\|_{n,\varpi}^2]/2 \leqslant C\bar f s/n.\end{aligned}$$

Thus, by Markov's inequality we have $P(\widehat{R}_u(\bar\beta_u) \leqslant C\bar f s/n) \geqslant 1/2$.

Define $z_{ui} := -\int_0^1 1\{\epsilon_{ui} \leqslant -tr_{ui}\} - 1\{\epsilon_{ui} \leqslant 0\}\, dt$, so that $\widehat{R}_u(\bar\beta_u) = \mathbb{E}_n[K_\varpi(W) r_u z_u]$ where $|z_{ui}| \leqslant 1$. We have $P(\mathbb{E}_n[K_\varpi(W) r_u z_u] \leqslant 2C\bar f s/n) \geqslant 1/2$ so that by Lemma 2.3.7 in [62] (note that the Lemma does not require zero mean stochastic processes), for $t \geqslant 2C\bar f s/n$ we have

$$\frac{1}{2}P(\sup_{u\in\mathcal{U}} |\mathbb{E}_n[K_\varpi(W) r_u z_u]| \geqslant t) \leqslant 2P(\sup_{u\in\mathcal{U}} |\mathbb{E}_n[K_\varpi(W) r_u z_u \epsilon]| > t/4)$$

Consider the class of functions $\mathcal{F} = \{-K_\varpi(W) r_u(1\{\epsilon_{ui} \leqslant -B_i r_{ui}\} - 1\{\epsilon_{ui} \leqslant 0\}) : u \in \mathcal{U}\}$ where $B_i \sim \text{Uniform}(0,1)$ independent of $(X_i, W_i)_{i=1}^n$. It follows that $K_\varpi(W) r_u z_u = \mathrm{E}[-K_\varpi(W) r_u(1\{\epsilon_{ui} \leqslant -B_i r_{ui}\} - 1\{\epsilon_{ui} \leqslant 0\}) \mid X_i, W_i]$ where the expectation is taken over B_i only. Thus we will bound the entropy of $\overline{\mathcal{F}} = \{\mathrm{E}[f \mid X, W] : f \in \mathcal{F}\}$ via Lemma 24. Note that $\mathcal{R} := \{r_u = X_{-a}\beta_u - X_{-a}\bar\beta_u : u \in \mathcal{U}\}$ where $\mathcal{G} := \{X_{-a}\bar\beta_u : u \in \mathcal{U}\}$ is contained in the union of at most $|V|\binom{p}{s}$ VC-classes of dimension Cs and $\mathcal{H} := \{X_{-a}\beta_u : u \in \mathcal{U}\}\}$ is a VC-class of functions of dimension $(1+d_W)$ by Condition P. Finally note that $\mathcal{E} := \{\epsilon_{ui} : u \in \mathcal{U}\} \subset \{X_{ai} : a \in V\} - \mathcal{G} - \mathcal{R}$.

Therefore, we have

$$\begin{aligned}\sup_Q \log N(\epsilon\|\bar F\|_{Q,2}, \overline{\mathcal{F}}, \|\cdot\|_{Q,2}) &\leqslant \sup_Q \log N((\epsilon/4)^2\|F\|_{Q,2}, \mathcal{F}, \|\cdot\|_{Q,2})\\ &\leqslant \sup_Q \log N(\tfrac{1}{8}(\epsilon^2/16), \mathcal{W}, \|\cdot\|_{Q,2})\\ &+ \sup_Q \log N(\tfrac{1}{8}(\epsilon^2/16)\|F\|_{Q,2}, \mathcal{R}, \|\cdot\|_{Q,2})\\ &+ \sup_Q \log N(\tfrac{1}{8}(\epsilon^2/16), 1\{\mathcal{E} + \{B\}\mathcal{R} \leqslant 0\} - 1\{\mathcal{E} \leqslant 0\}, \|\cdot\|_{Q,2})\end{aligned}$$

By Lemma 18 with envelope $\bar F = \|X\|_\infty \sup_{u\in\mathcal{U}} \|\beta_u - \bar\beta_u\|_1$, and $\sigma^2 = \sup_{u\in\mathcal{U}} \mathrm{E}[K_\varpi(W) r_{ui}^2] \lesssim s/n$ by Condition P, we have that with probability $1-o(1)$

$$\sup_{u\in\mathcal{U}} |\mathbb{E}_n[K_\varpi(W) r_u z_u \epsilon]| \lesssim \sqrt{\frac{s(1+d_W)\log(p|V|n)}{n}}\sqrt{\frac{s}{n}} + \frac{M_n\sqrt{s^2/n}\log(p|V|n)}{n} \lesssim \frac{s(1+d_W)\log(p|V|n)}{n}$$

under $M_n\sqrt{s^2/n} \leqslant C$. ∎

Lemma 10 (PQGM, Event Ω_3)**.** *Under Condition P, for* $u = (a,\tau,\varpi) \in V\times\mathcal{T}\times\mathcal{W}$*, define* $g_u(\delta, X, W) = K_\varpi(W)\{\rho_\tau(X_a - X_{-a}(\beta_u+\delta)) - \rho_\tau(X_a - X_{-a}\beta_u)\}$*, and*

$$\Omega_3 := \left\{ \sup_{u\in\mathcal{U}, 1/\sqrt{n}\leqslant\|\delta\|_{1,\varpi}\leqslant\sqrt{n}} \frac{|\mathbb{E}_n[g_u(\delta,X,W) - \mathrm{E}[g_u(\delta,X,W) \mid X_{-a}]]|}{\|\delta\|_{1,\varpi}} < t_3 \right\}.$$

Then, under Condition CI we have $P(\Omega_3) \geqslant 1-\gamma$ for any

$$t_3\sqrt{n} \geqslant 12 + 16\sqrt{2\log\left(64|V|p^2 n^{1+d_W}\log(n) L_{\mathcal{WT}}^{1+d_W} M_n^{1+d_W/\rho}/\gamma\right)}$$

Proof. We have that $\Omega_3^c := \{\max_{a\in V} A_a \geqslant t_3\sqrt{n}\}$ for

$$A_a := \sup_{(\tau,\varpi)\in\mathcal{T}\times\mathcal{W}, \underline{N}\leqslant\|\delta\|_{1,\varpi}\leqslant\bar{N}} \sqrt{n}\left|\frac{\mathbb{E}_n[g_u(\delta,X,W) - \mathrm{E}[g_u(\delta,X,W) \mid X_{-a},W]]}{\|\delta\|_{1,\varpi}}\right|.$$

Therefore, for $\underline{N} = 1/\sqrt{n}$ and $\bar{N} = \sqrt{n}$ we have by Lemma 13

$$\begin{aligned} P(\Omega_3^c) &= P(\max_{a\in V} A_a \geqslant t_3\sqrt{n}) \\ &\leqslant |V|\max_{a\in V} P(A_a \geqslant t_3\sqrt{n}) \\ &= |V|\max_{a\in V} \mathrm{E}_{X_{-a},W}\{P(A_a \geqslant t_3\sqrt{n} \mid X_{-a},W)\} \\ &\leqslant |V|\max_{a\in V} \mathrm{E}_{X_{-a},W}\left\{8p|\widehat{\mathcal{N}}|\cdot|\widehat{\mathcal{W}}|\cdot|\widehat{\mathcal{T}}|\exp(-(t_3\sqrt{n}/4-3)^2/32)\right\} \\ &\leqslant \exp(-(t_3\sqrt{n}/4-3)^2/32)|V|64pn^{1+d_W}\log(n)L_f \mathrm{E}_{X_{-a}}\left\{\frac{\max_{i\leqslant n}\|X_{-ai}\|_\infty^{1+d_W}}{\underline{N}^{1+d_W/\rho}}\right\} \\ &\leqslant \gamma \end{aligned}$$

by the choice of t_3. ∎

Lemma 11 (PQGM, Event Ω_4). *Under Condition P, and setting $K_u = C\sqrt{\frac{(1+d_W)\log(p|V|n)}{n}}$, we have that* $\mathrm{P}(\Omega_4) = 1 - o(1)$.

Proof. First note that by Lemma 18 we have that with probability $1-o(1)$

$$\sup_{u\in\mathcal{U},j\in V}(\mathbb{E}_n - \mathrm{E})[K_\varpi(W)X_j^2] \lesssim \sqrt{\frac{(1+d_W)\log(|V|M_n/\sigma)}{n}} + \frac{(1+d_W)M_n^2\log(|V|M_n/\sigma)}{n}$$

and $\mathrm{E}[K_\varpi(W)X_j^2] \geqslant c\mathrm{P}(\varpi)$. Under $(1+d_W)\log(|V|M_n/\sigma) \leqslant \delta_n^2\mu_{\mathcal{W}}^2$ and $(1+d_W)M_n^2\log(|V|M_n/\sigma) \leqslant \delta_n n\mu_{\mathcal{W}}$, we have that $|(\mathbb{E}_n - \mathrm{E})[K_\varpi(W)X_j^2]| = o(\mathrm{E}[K_\varpi(W)X_j^2])$ for all $u\in\mathcal{U}$. Therefore, we have

$$\begin{aligned} &\mathrm{P}(\sup_{u\in\mathcal{U}}|\mathbb{E}_n[h_{uj}(X_{-a},W)]|/\{K_u\mathbb{E}_n[K_\varpi(W)X_j^2]^{1/2}\} > 1) \\ &\leqslant o(1) + \mathrm{P}(\sup_{u\in\mathcal{U}}|\mathbb{E}_n[h_{uj}(X_{-a},W)]|/\{K_u\mathrm{E}[K_\varpi(W)X_j^2]^{1/2}\} > 1 + O(\delta_n)) \end{aligned}$$

Applying Lemma 18 to $\mathcal{F} = \{h_{uj}(X_{-a},W)/\mathrm{E}[K_\varpi(W)X_j^2]^{1/2} : u\in\mathcal{U}\}$. For convenience define $\bar{\mathcal{H}}_j = \{h_{uj}(X_{-a},W) : u\in\mathcal{U}\}$ and $\bar{\mathcal{K}}_j := \{\mathrm{E}[K_\varpi(W)X_j^2] : \varpi\in\mathcal{W}\}$. Note that $\bar{\mathcal{K}}_j$ has covering numbers bounded by the covering number of $\mathcal{K}_j := \{K_\varpi(W)X_j^2 : \varpi\in\mathcal{W}\}$ as follows $\sup_Q \log N(\epsilon\|\bar{K}_j\|_{Q,2}, \bar{\mathcal{K}}_j, \|\cdot\|_{Q,2}) \leqslant \log\sup_{\tilde{Q}} N((\epsilon/4)^2\|F\|_{\tilde{Q},2}, \mathcal{K}_j, \|\cdot\|_{\tilde{Q},2})$ by Lemma 24. Similarly, Lemma 24 also allows us to bound covering numbers of $\bar{\mathcal{H}}_j$ via covering numbers of $\mathcal{H}_j = \{K_\varpi(W)(\tau - 1\{X_a \leqslant X_{-a}\beta_u\})X_j : u\in\mathcal{U}\}$.

$$\begin{aligned} &\sup_Q \log N(\epsilon\|F\|_{Q,2}, \mathcal{F}, \|\cdot\|_{Q,2}) \leqslant p\max_{j\in[p]}\sup_Q \log N(\epsilon\|F_j\|_{Q,2}, \mathcal{F}_j, \|\cdot\|_{Q,2}) \\ &\leqslant p\max_{j\in[p]}\sup_Q\{\log N((1/2)\epsilon\|\bar{H}_j\|_{Q,2}, \bar{\mathcal{H}}_j, \|\cdot\|_{Q,2}) + \sup_Q \log N((1/2)\epsilon c\mu_{\mathcal{W}}^{1/2}, 1/\bar{\mathcal{K}}_j^{1/2}, \|\cdot\|_{Q,2})\} \\ &\leqslant p\max_{j\in[p]}\sup_Q\{\log N((1/2)\epsilon\|\bar{H}_j\|_{Q,2}, \bar{\mathcal{H}}_j, \|\cdot\|_{Q,2}) + \sup_Q \log N((1/2)\epsilon c\mu_{\mathcal{W}}^{3/2}, \bar{\mathcal{K}}_j^{1/2}, \|\cdot\|_{Q,2})\} \\ &\leqslant p\max_{j\in[p]}\sup_Q\{\log N((1/2)\epsilon\|\bar{H}_j\|_{Q,2}, \bar{\mathcal{H}}_j, \|\cdot\|_{Q,2}) + \sup_Q \log N(C(1/2)\epsilon c\mu_{\mathcal{W}}^{3/2}, \bar{\mathcal{K}}_j, \|\cdot\|_{Q,2})\} \\ &\leqslant p\max_{j\in[p]}\sup_Q\{\log N((1/2)\epsilon\|\bar{H}_j\|_{Q,2}, \bar{\mathcal{H}}_j, \|\cdot\|_{Q,2}) + \sup_Q \log N((1/\{2C\})\epsilon c\mu_{\mathcal{W}}^{3/2}, \bar{\mathcal{K}}_j, \|\cdot\|_{Q,2})\} \\ &\leqslant p\max_{j\in[p]}\sup_{\tilde{Q}} \log N((1/4)\epsilon^2\|H_j\|_{Q,2}, \mathcal{H}_j, \|\cdot\|_{Q,2}) \\ &+ p\max_{j\in[p]}\sup_{\tilde{Q}} \log N((1/\{4C^2\})\epsilon^2 c^2\mu_{\mathcal{W}}^3, \mathcal{K}_j, \|\cdot\|_{\tilde{Q},2}) \end{aligned}$$

where $F_j = c\|X\|_\infty/\mu_{\mathcal{W}}^{1/2}$, $H_j = \|X\|_\infty$. Since $\mathcal{K}_j$ is the product of a VC subgraph of dimension d_W with a single function, and $\mathcal{H}_j$ is the product of two VC subgraph of dimension $1+d_W$ and a single function, by Lemma 18 with $\sigma^2 = 1$, we have with probability $1-o(1)$

$$\sup_{u\in\mathcal{U}}\left|\frac{(\mathbb{E}_n - \mathrm{E})[h_{uj}(X_{-a},W)]}{\mathrm{E}[K_\varpi(W)X_j^2]^{1/2}}\right| \leqslant C\sqrt{\frac{(1+d_W)\log(p|V|n)}{n}} + C\frac{M_n(1+d_W)\log(p|V|n)}{n\mu_{\mathcal{W}}^{1/2}}.$$

Thus, under $M_n(1+d_W)\log(p|V|n) \leqslant n^{1/2}\mu_{\mathcal{W}}^{1/2}$ we have that we can take $K_u = C\sqrt{\frac{(1+d_W)\log(p|V|n)}{n}}$. ∎

Appendix F. Technical Results for High-Dimensional Quantile Regression

In this section we provide technical results for high-dimensional quantile regression. It is based on a sample $(\tilde{y}_i, \tilde{x}_i, W_i)_{i=1}^n$, independent across i, $\rho_\tau(t) = (\tau - 1\{t \leqslant 0\})t$, $\tau \in \mathcal{T} \subset (0,1)$ a compact interval, and a family of indicator functions $K_w(W) = 1$ if $W \in \Omega_\varpi$, where $\Omega_\varpi \in \mathcal{W}$. For convenience we index the sets Ω_ϖ by $\varpi \in B_W \subset \mathbb{R}^{d_W}$ where we normalize the diameter of B_W to be less or equal than 1/6. Let $f_{\tilde{y}|\tilde{x},r_u,\varpi}(\cdot)$ denote the conditional probability density function, $f_u := f_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u)$, $f_{\tilde{y}|\tilde{x},r_u,\varpi}(\cdot) \leqslant \bar{f}$ and $|f'_{\tilde{y}|\tilde{x},r_u,\varpi}(\cdot)| \leqslant \bar{f}'$. Moreover, we assume that $\|\eta_u - \eta_{\tilde{u}}\|_1 \leqslant L_f\{|\tau - \tilde{\tau}| + \|\varpi - \tilde{\varpi}\|^\rho\}$.

Although the results can be applied more generally, these results will be used for (η_u, r_u), $u = (\tilde{y}, \tau, \varpi) \in \{\tilde{y}\} \times \mathcal{T} \times \mathcal{W} \in \mathcal{U}$ satisfying

$$\mathrm{E}[K_\varpi(W)(\tau - 1\{\tilde{y} \leqslant \tilde{x}'\eta_u + r_u\})\tilde{x}] = 0.$$

Note that this generality is flexible enough to allow us to cover the case that the τ-conditional quantile function $Q_{\tilde{y}}(\tau \mid \tilde{x}_i, \varpi) = \tilde{x}_i'\tilde{\eta}_u + \tilde{r}_u$ by setting $\eta_u = \tilde{\eta}_u$ and $r_u = \tilde{r}_u$ in which case $\mathrm{E}[(\tau - 1\{\tilde{y} \leqslant \tilde{x}'\beta_u + r_u\}) \mid \tilde{x}, \varpi] = 0$. It also covers the case that

$$\tilde{\eta}_u \in \arg\min_\beta \mathrm{E}[K_\varpi(W)\rho_\tau(\tilde{y} - \tilde{x}'\beta)]$$

so that $\mathrm{E}[K_\varpi(W)(\tau - 1\{\tilde{y} \leqslant \tilde{x}'\tilde{\eta}_u\})\tilde{x}] = 0$ holds by the first order condition by setting $\eta_u = \tilde{\eta}_u$ and $r_u = 0$. Moreover, it also covers the case that we work with a sparse approximation $\bar{\eta}_u$ of $\tilde{\eta}_u$ by setting $\eta_u = \bar{\eta}_u$ and $r_u = \tilde{x}'(\tilde{\eta}_u - \bar{\eta}_u)$.

Lemma 12 (Identification Lemma). *For $u = (a,\tau,\varpi) \in \mathcal{U}$, and a subset $A_u \subset \mathbb{R}^p$ let*

$$\bar{q}_{A_u} = (1/2)\cdot(1/\bar{f}')\cdot \inf_{\delta\in A_u} \mathbb{E}_n\left[K_\varpi(W)f_u|\tilde{x}'\delta|^2\right]^{3/2} / \mathbb{E}_n\left[K_\varpi(W)|\tilde{x}'\delta|^3\right]$$

and assume that for all $\delta \in A_u$

$$\mathbb{E}_n\left[K_\varpi(W)|r_u|\cdot|\tilde{x}'\delta|^2\right] + \mathbb{E}_n\left[K_\varpi(W)r_u^2\cdot|\tilde{x}'\delta|^2\right] \leqslant (1/[4\bar{f}'])\mathbb{E}_n[K_\varpi(W)f_u|\tilde{x}'\delta|^2]. \tag{F.44}$$

Then we have

$$\begin{aligned}&\mathbb{E}_n\mathrm{E}[K_\varpi(W)\rho_\tau(\tilde{y} - \tilde{x}'(\eta_u + \delta)) \mid \tilde{x}, r_u, W] - \mathbb{E}_n\mathrm{E}[K_\varpi(W)\rho_\tau(\tilde{y} - \tilde{x}'\eta_u) \mid \tilde{x}, r_u, W]\\ &\geqslant \tfrac{\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi}^2}{4} \wedge \left\{\bar{q}_A\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi}\right\} - K_{n2}\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi} - K_{n1}\|\delta\|_{1,\varpi}.\end{aligned}$$

where $K_{n2} := (\bar{f}^{1/2}+1)\|r_u\|_{n,\varpi}$ and $K_{n1} := \sup_{u\in\mathcal{U},j\in[p]} \frac{|\mathbb{E}_n\mathrm{E}[K_\varpi(W)(\tau - 1\{\tilde{y}\leqslant\tilde{x}'\eta_u + r_u\})\tilde{x}_j|\tilde{x},W]|}{\{\mathbb{E}_n[K_\varpi(W)\tilde{x}_j^2]\}^{1/2}}$.

Proof of Lemma 12. Let $T_u = \text{support}(\eta_u)$, and $Q_u(\eta) := \bar{\mathrm{E}}[K_\varpi(W)\rho_\tau(\tilde{y} - \tilde{x}'\eta) \mid \tilde{x}, r_u, W]$. The proof proceeds in steps.

Step 1. (Minoration). Define the maximal radius over which the criterion function can be minorated by a quadratic function

$$r_{A_u} = \sup_r \left\{ \begin{array}{rl} r \ : & Q_u(\eta_u+\delta) - Q_u(\eta_u) + K_{n2}\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi} + K_{n1}\|\delta\|_{1,\varpi} \geqslant \frac{1}{4}\|\sqrt{f_u}\tilde{x}'\delta\|^2_{n,\varpi}, \\ & \forall \delta \in A_u, \ \|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi} \leqslant r \end{array} \right\}.$$

Step 2 below shows that $r_{A_u} \geqslant \bar{q}_{A_u}$. By construction of r_{A_u} and the convexity of $Q_u(\cdot)$, $\|\cdot\|_{1,\varpi}$ and $\|\cdot\|_{n,\varpi}$,

$$\begin{array}{l} Q_u(\eta_u+\delta) - Q_u(\eta_u) + K_{n2}\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi} + K_{n1}\|\delta\|_{1,\varpi} \geqslant \\ \geqslant \frac{\|\sqrt{f_u}\tilde{x}'\delta\|^2_{n,\varpi}}{4} \wedge \left\{ \frac{\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi}}{r_A} \cdot \begin{array}{rl} \inf & Q_u(\eta_u+\tilde{\delta}) - Q_u(\eta_u) + K_{n2}\|\sqrt{f_u}\tilde{x}'\tilde{\delta}\|_{n,\varpi} + K_{n1}\|\tilde{\delta}\|_{1,\varpi} \\ & \tilde{\delta} \in A_u, \|\sqrt{f_u}\tilde{x}'\tilde{\delta}\|_u \geqslant r_{A_u} \end{array} \right\} \\ \geqslant \frac{\|\sqrt{f_u}\tilde{x}'\delta\|^2_{n,\varpi}}{4} \wedge \left\{ \frac{\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi}}{r_{A_u}} \frac{r^2_{A_u}}{4} \right\} \\ \geqslant \frac{\|\sqrt{f_u}\tilde{x}'\delta\|^2_{n,\varpi}}{4} \wedge \left\{ \bar{q}_{A_u}\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi} \right\}. \end{array}$$

Step 2. ($r_{A_u} \geqslant \bar{q}_{A_u}$) Let $F_{\tilde{y}|\tilde{x}}$ denote the conditional distribution of $\tilde{y}$ given $\tilde{x}$. From [41], for any two scalars w and v the Knight's identity is

$$\rho_\tau(w-v) - \rho_\tau(w) = -v(\tau - 1\{w \leqslant 0\}) + \int_0^v (1\{w \leqslant z\} - 1\{w \leqslant 0\})dz. \tag{F.45}$$

Using (F.45) with $w = \tilde{y}_i - \tilde{x}_i'\eta_u$ and $v = \tilde{x}_i'\delta$ and taking expectations with respect to $\tilde{y}$, we have

$$\begin{array}{rl} Q_u(\eta_u+\delta) - Q_u(\eta_u) = & -\mathbb{E}_n \mathrm{E}\left[K_\varpi(W)(\tau - 1\{\tilde{y} \leqslant \tilde{x}'\eta_u\})\tilde{x}_i'\delta \mid \tilde{x}, r_u, W\right] \\ & +\mathbb{E}_n \left[\int_0^{K_\varpi(W)\tilde{x}'\delta} F_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u + t) - F_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u)dt\right]. \end{array}$$

Using the law of iterated expectations and mean value expansion, the relation

$$\begin{array}{l} |\mathbb{E}_n \mathrm{E}\left[K_\varpi(W)(\tau - 1\{\tilde{y} \leqslant \tilde{x}'\eta_u\})\tilde{x}'\delta \mid \tilde{x}, r_u, W\right]| \\ = |\mathbb{E}_n \left[K_\varpi(W)\{F_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u + r_u) - F_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u)\}\tilde{x}'\delta\right] \\ +\mathbb{E}_n \left[K_\varpi(W)\{\tau - F_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u + r_u)\}\tilde{x}'\delta \mid \tilde{x}, r_u, W\right]| \\ \leqslant \mathbb{E}_n[K_\varpi(W)f_u|r_u| \ |\tilde{x}'\delta|] + \bar{f}'\mathbb{E}_n[K_\varpi(W)|r_u|^2|\tilde{x}'\delta|] + K_{n1}\|\delta\|_{1,\varpi} \\ \leqslant \|\sqrt{f_u}r_u\|_{n,\varpi}\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi} + \bar{f}'\|r_u\|_{n,\varpi}\|r_u\tilde{x}'\delta\|_{n,\varpi} + K_{n1}\|\delta\|_{1,\varpi} \\ \leqslant (\bar{f}^{1/2}+1)\|r_u\|_{n,\varpi}\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi} + K_{n1}\|\delta\|_{1,\varpi} \end{array}$$

where we used our assumption on the approximation error and we have $K_{n2} = (\bar{f}^{1/2}+1)\|r_u\|_{n,\varpi}$. With that and similar arguments we obtain for $\tilde{t}_{\tilde{x}_i,t} \in [0,t]$

$$\begin{array}{l} Q_u(\eta_u+\delta) - Q_u(\eta_u) + K_{n2}\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi} + K_{n1}\|\delta\|_{1,\varpi} \geqslant \\ Q_u(\eta_u+\delta) - Q_u(\eta_u) + \mathbb{E}_n \mathrm{E}\left[K_\varpi(W)(\tau - 1\{\tilde{y} \leqslant \tilde{x}'\eta_u\})\tilde{x}'\delta \mid \tilde{x}, r_u, W\right] = \\ = \mathbb{E}_n \left[\int_0^{K_\varpi(W)\tilde{x}'\delta} F_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u + t) - F_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u)dt\right] \\ = \mathbb{E}_n \left[\int_0^{K_\varpi(W)\tilde{x}'\delta} t f_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u) + \frac{t^2}{2} f'_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u + \tilde{t}_{\tilde{x},t})dt\right] \\ \geqslant \frac{1}{2}\|\sqrt{f_u}\tilde{x}'\delta\|^2_{n,\varpi} - \frac{1}{6}\bar{f}'\mathbb{E}_n[K_\varpi(W)|\tilde{x}'\delta|^3] - \bar{\mathrm{E}}\left[\int_0^{K_\varpi(W)\tilde{x}_i'\delta} t[f_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u) - f_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u + r_u)]dt\right] \\ \geqslant \frac{1}{4}\|\sqrt{f_u}\tilde{x}'\delta\|^2_{n,\varpi} + \frac{1}{4}\|\sqrt{f_u}\tilde{x}'\delta\|^2_{n,\varpi} - \frac{1}{6}\bar{f}'\mathbb{E}_n[K_\varpi(W)|\tilde{x}'\delta|^3] - (\bar{f}'/2)\mathbb{E}_n\left[K_\varpi(W)|\tilde{r}_u| \cdot |\tilde{x}'\delta|^2\right]. \end{array} \tag{F.46}$$

Moreover, by assumption we have

$$\mathbb{E}_n\left[K_\varpi(W)|r_u|\cdot|\tilde{x}'\delta|^2\right] \leqslant \tfrac{1}{4\bar{f}'}\mathbb{E}_n[K_\varpi(W)f_u|\tilde{x}'\delta|^2] \tag{F.47}$$

Note that for any δ such that $\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi} \leqslant \bar{q}_A$ we have

$$\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi} \leqslant \bar{q}_{uA} \leqslant (1/2)\cdot(1/\bar{f}')\cdot\mathbb{E}_n\left[K_\varpi(W)f_u|\tilde{x}'\delta|^2\right]^{3/2}/\mathbb{E}_n\left[K_\varpi(W)|\tilde{x}'\delta|^3\right].$$

It follows that $(1/6)\bar{f}'\mathbb{E}_n[K_\varpi(W)|\tilde{x}'\delta|^3] \leqslant (1/8)\mathbb{E}_n[K_\varpi(W)f_u|\tilde{x}'\delta|^2]$. Combining this with (F.47) we have

$$\frac{1}{4}\mathbb{E}_n[K_\varpi(W)f_u|\tilde{x}'\delta|^2] - \frac{\bar{f}'}{6}\mathbb{E}_n[K_\varpi(W)|\tilde{x}'\delta|^3] - \frac{\bar{f}'}{2}\mathbb{E}_n\left[K_\varpi(W)|r_u|\cdot|\tilde{x}'\delta|^2\right] \geqslant 0. \tag{F.48}$$

Combining (F.46) and (F.48) we have $r_{A_u} \geqslant \bar{q}_{A_u}$. ∎

Lemma 13. *Let $\mathcal{W}$ be a VC-class of sets with VC-index d_W. Conditional on $\{(W_i,\tilde{x}_i), i=1,\ldots,n\}$ we have*

$$P_{\tilde{y}}\left(\sup_{\substack{\tau\in\mathcal{T},\varpi\in\mathcal{W},\\ \underline{N}\leqslant\|\delta\|_{1,\varpi}\leqslant\bar{N}}}\left|\mathbb{G}_n\left(K_\varpi(W)\frac{\rho_\tau(\tilde{y}-\tilde{x}'(\eta_u+\delta))-\rho_\tau(\tilde{y}-\tilde{x}'\eta_u)}{\|\delta\|_{1,\varpi}}\right)\right| \geqslant M \mid (W_i,\tilde{x}_i)_{i=1}^n\right) \leqslant S_n\exp(-(M/4-3)^2/32)$$

where $S_n \leqslant 8p|\widehat{\mathcal{N}}|\cdot|\widehat{\mathcal{W}}|\cdot|\widehat{\mathcal{T}}|$, with

$$|\widehat{\mathcal{N}}| \leqslant 1+\left\lfloor 3\sqrt{n}\log(\bar{N}/\underline{N})\right\rfloor, \quad |\widehat{\mathcal{T}}| \leqslant 2\sqrt{n}\frac{\max_{i\leqslant n}\|\tilde{x}_i\|_\infty}{\underline{N}}L_f, \quad |\widehat{\mathcal{W}}| \leqslant n^{d_W}+\left\{2\sqrt{n}\frac{\max_{i\leqslant n}\|\tilde{x}_i\|_\infty}{\underline{N}}L_f\right\}^{d_W/\rho}.$$

Proof of Lemma 13. Let $g_{\tau\varpi,i}(b) = K_\varpi(W_i)\{\rho_\tau(\tilde{y}_i-\tilde{x}_i'\eta_{\tau\varpi}+b)-\rho_\tau(\tilde{y}_i-\tilde{x}_i'\eta_{\tau\varpi})\} \leqslant K_\varpi(W_i)|b|$ since $K_\varpi(W_i)\in\{0,1\}$. Note that $|g_{\tau\varpi,i}(b)-g_{\tau\varpi,i}(a)| \leqslant K_\varpi(W_i)|b-a|$. To easy the notation we omit the conditioning on $(\tilde{x}_i,W_i)$ from the probabilities.

For any $\delta\in\mathbb{R}^p$, since ρ_τ is 1-Lipschitz, we have

$$\operatorname{var}\left(\mathbb{G}_n\left(\tfrac{g_{\tau\varpi}(\tilde{x}'\delta)}{\|\delta\|_{1,\varpi}}\right)\right) \leqslant \tfrac{\mathbb{E}_n[\{g_{\tau\varpi}(\tilde{x}'\delta)\}^2]}{\|\tilde{x}'\delta\|^2_{n,\varpi}} \leqslant \tfrac{\mathbb{E}_n[|K_\varpi(W_i)\tilde{x}'\delta|^2]}{\|\tilde{x}'\delta\|^2_{n,\varpi}} = 1$$

since by definition $\|\delta\|_{1,\varpi} = \sum_j\|\delta_j\|_{1,\varpi} = \sum_j\|\tilde{x}'\delta_j\|_{n,\varpi} \geqslant \|\tilde{x}'\delta\|_{n,\varpi}$.

Since we are conditioning on $(W_i,\tilde{x}_i)_{i=1}^n$ the processes are independent across i. Then, by Lemma 2.3.7 in [61] (Symmetrization for Probabilities) we have for any $M>1$

$$\begin{aligned}&\mathrm{P}\left(\sup_{\tau\in\mathcal{T},\varpi\in\mathcal{W},\underline{N}\leqslant\|\delta\|_{1,\varpi}\leqslant\bar{N}}\left|\mathbb{G}_n\left(\frac{g_{\tau\varpi}(\tilde{x}'\delta)}{\|\delta\|_{1,\varpi}}\right)\right| \geqslant M\right)\\ &\leqslant \frac{2}{1-M^{-2}}P\left(\sup_{\tau\in\mathcal{T},\varpi\in\mathcal{W},\underline{N}\leqslant\|\delta\|_{1,\varpi}\leqslant\bar{N}}\left|\mathbb{G}_n^o\left(\frac{\kappa_{a\tau\varpi,\mathbf{c}}g_{\tau\varpi}(\tilde{x}'\delta)}{\|\delta\|_{1,\varpi}}\right)\right| \geqslant M/4\right)\end{aligned}$$

where $\mathbb{G}_n^o$ is the symmetrized process.

Consider $\mathcal{F}_{t,\tau,\varpi} = \{\delta : \|\delta\|_{1,\varpi} = t\}$. We will consider the families of $\mathcal{F}_{t,\tau,\varpi}$ for $t\in[\underline{N},\bar{N}]$, $\tau\in\mathcal{T}$ and $\varpi\in\mathcal{W}$.

We will construct a finite net $\widehat{\mathcal{T}}\times\widehat{\mathcal{W}}\times\widehat{\mathcal{N}}$ of $\mathcal{T}\times\mathcal{W}\times[\underline{N},\bar{N}]$ such that

$$\sup_{\tau\in\mathcal{T},\varpi\in\mathcal{W},t\in[\underline{N},\bar{N}],\delta\in\mathcal{F}_{t,\tau,\varpi}}\left|\mathbb{G}_n^o\left(\frac{g_{\tau\varpi}(\tilde{x}'\delta)}{\|\delta\|_{1,\varpi}}\right)\right| \leqslant 3+\sup_{\tau\in\widehat{\mathcal{T}},\varpi\in\widehat{\mathcal{W}},t\in\widehat{\mathcal{N}}}\sup_{\delta\in\mathcal{F}_{t,\tau,\varpi}}\left|\mathbb{G}_n^o\left(\frac{g_{\tau\varpi}(\tilde{x}'\delta)}{t}\right)\right| =: 3+\mathcal{A}^o.$$

By triangle inequality we have

$$\begin{aligned}\left|\mathbb{G}_n^o\left(\frac{g_{\tau\varpi}(\tilde{x}'\delta)}{t}-\frac{g_{\tilde{\tau}\tilde{\varpi}}(\tilde{x}_i'\tilde{\delta})}{\tilde{t}}\right)\right| &\leqslant \left|\mathbb{G}_n^o\left(\frac{g_{\tau\varpi,i}(\tilde{x}_i'\delta)}{t}-\frac{g_{\tilde{\tau}\varpi}(\tilde{x}'\delta)}{t}\right)\right|+\left|\mathbb{G}_n^o\left(\frac{g_{\tilde{\tau}\varpi}(\tilde{x}'\delta)}{t}-\frac{g_{\tilde{\tau}\tilde{\varpi}}(\tilde{x}'\delta)}{t}\right)\right| \\ &\quad+\left|\mathbb{G}_n^o\left(\frac{w_{\tilde{\tau}\tilde{\varpi}}(\tilde{x}'\delta)}{t}-\frac{g_{\tilde{\tau}\tilde{\varpi}}(\tilde{x}'\tilde{\delta})}{\tilde{t}}\right)\right|\end{aligned} \tag{F.49}$$

The first term in (F.49) is such that

$$\begin{aligned}\left|\mathbb{G}_n^o\left(\frac{g_{\tau\varpi}(\tilde{x}'\delta)}{t}-\frac{g_{\tilde{\tau}\varpi}(\tilde{x}'\delta)}{t}\right)\right| &\leqslant \frac{2\sqrt{n}}{t}\mathbb{E}_n[K_\varpi(W)|\tilde{x}'(\eta_{\tau\varpi}-\eta_{\tilde{\tau}\varpi})|] \\ &\leqslant \frac{2\sqrt{n}}{\underline{N}}\max_{i\leqslant n}\|\tilde{x}_i\|_\infty\mathbb{E}_n[K_\varpi(W)]\|\eta_{\tau\varpi}-\eta_{\tilde{\tau}\varpi}\|_1 \\ &\leqslant \frac{2\sqrt{n}}{\underline{N}}\max_{i\leqslant n}\|\tilde{x}_i\|_\infty\mathbb{E}_n[K_\varpi(W)]L_f|\tau-\tau'|.\end{aligned} \tag{F.50}$$

Define a net $\widehat{\mathcal{T}}=\{\tau_1,\ldots,\tau_T\}$ such that

$$|\tau_{k+1}-\tau_k|\leqslant\left\{2\sqrt{n}\frac{\max_{i\leqslant n}\|\tilde{x}_i\|_\infty}{\underline{N}}L_f\right\}^{-1}$$

To bound the second term in (F.49), note that $\mathcal{W}$ is a VC-class. Therefore, by Corollary 2.6.3 in [61] we have that conditionally on $(W_i)_{i=1}^n$, there are at most n^{d_W} different sets $\varpi\in\mathcal{W}$ that induce a different sequence $\{K_\varpi(W_1),\ldots,K_\varpi(W_n)\}$. We that a cover $\widehat{\mathcal{W}}$ that covers all different sequences. Further, similarly to (F.51) we have $\|\eta_{\tilde{\tau}\varpi}-\eta_{\tilde{\tau}\tilde{\varpi}}\|_1\leqslant L_f\|\varpi-\tilde{\varpi}\|^\rho$ and

$$\begin{aligned}\left|\mathbb{G}_n^o\left(\frac{g_{\tilde{\tau}\varpi}(\tilde{x}'\delta)}{t}-\frac{g_{\tilde{\tau}\tilde{\varpi}}(\tilde{x}'\delta)}{t}\right)\right| &\leqslant \frac{2\sqrt{n}}{t}\mathbb{E}_n[K_\varpi(W)|\tilde{x}'(\eta_{\tilde{\tau}\varpi}-\eta_{\tilde{\tau}\tilde{\varpi}})|]+\frac{2\sqrt{n}}{t}\mathbb{E}_n[|K_\varpi(W)-K_{\tilde{\varpi}}(W)||\tilde{x}'\delta|] \\ &\leqslant \frac{2\sqrt{n}}{\underline{N}}\max_{i\leqslant n}\|\tilde{x}_i\|_\infty\mathbb{E}_n[K_\varpi(W)]\|\eta_{\tilde{\tau}\varpi}-\eta_{\tilde{\tau}\tilde{\varpi}}\|_1 \\ &\leqslant \frac{2\sqrt{n}}{\underline{N}}\max_{i\leqslant n}\|\tilde{x}_i\|_\infty\mathbb{E}_n[K_\varpi(W)]L_f\|\varpi-\tilde{\varpi}\|^\rho.\end{aligned} \tag{F.51}$$

We define a net $\widehat{\mathcal{W}}$ such that $|\widehat{\mathcal{W}}|\leqslant n^{d_W}+\left\{2\sqrt{n}\frac{\max_{i\leqslant n}\|\tilde{x}_i\|_\infty}{\underline{N}}L_f\right\}^{d_W/\rho}$

To bound the third term, for any $\delta\in\mathcal{F}_{t,\tau,\varpi}$, $t\leqslant\tilde{t}$, we will choose $\tilde{\delta}:=\delta(\tilde{t}/t)\in\mathcal{F}_{\tilde{t},\tau,\varpi}$

$$\begin{aligned}\left|\mathbb{G}_n^o\left(\frac{g_{\tau\varpi}(\tilde{x}'\delta)}{t}-\frac{g_{\tau\varpi}(\tilde{x}'\delta(\tilde{t}/t))}{\tilde{t}}\right)\right| &\leqslant \left|\mathbb{G}_n^o\left(\frac{g_{\tau\varpi}(\tilde{x}'\delta)}{t}-\frac{g_{\tau\varpi}(\tilde{x}'\delta(\tilde{t}/t))}{t}\right)\right|+\left|\mathbb{G}_n^o\left(\frac{g_{\tau\varpi}(\tilde{x}'\delta(\tilde{t}/t))}{t}-\frac{g_{\tau\varpi}(\tilde{x}'\delta(\tilde{t}/t))}{\tilde{t}}\right)\right| \\ &= \frac{1}{t}\left|\mathbb{G}_n^o\left(g_{\tau\varpi}(\tilde{x}'\delta)-g_{\tau\varpi}(\tilde{x}'\delta[\tilde{t}/t])\right)\right|+\left|\mathbb{G}_n^o\left(g_{\tau\varpi}(\tilde{x}'\delta(\tilde{t}/t))\right)\right|\cdot\left|\frac{1}{t}-\frac{1}{\tilde{t}}\right| \\ &\leqslant \sqrt{n}\mathbb{E}_n\left(\frac{|K_\varpi(W)\tilde{x}'\delta|}{t}\right)\frac{|t-\tilde{t}|}{t}+\sqrt{n}\mathbb{E}_n\left(|K_\varpi(W)\tilde{x}'\delta|\right)\frac{\tilde{t}}{t}\left|\frac{1}{t}-\frac{1}{\tilde{t}}\right| \\ &= 2\sqrt{n}\mathbb{E}_n\left(\frac{|K_\varpi(W)\tilde{x}'\delta|}{t}\right)\left|\frac{t-\tilde{t}}{t}\right|\leqslant 2\sqrt{n}\left|\frac{t-\tilde{t}}{t}\right|.\end{aligned}$$

Let $\widehat{\mathcal{N}}$ be a ε-net $\{\underline{N}=:t_1,t_2,\ldots,t_K:=\bar{N}\}$ of $[\underline{N},\bar{N}]$ such that $|t_k-t_{k+1}|/t_k\leqslant 1/[2\sqrt{n}]$. Note that we can achieve that with $|\widehat{\mathcal{N}}|\leqslant 1+\lfloor 3\sqrt{n}\log(\bar{N}/\underline{N})\rfloor$.

By Markov bound, we have

$$\begin{aligned}P(\mathcal{A}^o\geqslant K) &\leqslant \min_{\psi\geqslant 0}\exp(-\psi K)\mathrm{E}[\exp(\psi\mathcal{A}^o)] \\ &\leqslant 8p|\widehat{\mathcal{T}}|\cdot|\widehat{\mathcal{W}}|\cdot|\widehat{\mathcal{N}}|\min_{\psi\geqslant 0}\exp(-\psi K)\exp\left(8\psi^2\right) \\ &\leqslant 8p|\widehat{\mathcal{T}}|\cdot|\widehat{\mathcal{W}}|\cdot|\widehat{\mathcal{N}}|\exp(-K^2/32)\end{aligned}$$

where we set $\psi = K/16$ and bound $\mathrm{E}[\exp(\psi\mathcal{A}^o)]$ as follows

$$\begin{aligned}
\mathrm{E}\left[\exp\left(\psi\mathcal{A}^o\right)\right] &\leqslant_{(1)} 2|\widehat{\mathcal{T}}|\cdot|\widehat{\mathcal{W}}|\cdot|\widehat{\mathcal{N}}| \sup_{(\tau,\varpi,t)\in\widehat{\mathcal{T}}\times\widehat{\mathcal{W}}\times\widehat{\mathcal{N}}} \mathrm{E}\left[\exp\left(\psi \sup_{\|\delta\|_{1,\varpi}=t} \mathbb{G}_n^o\left(\frac{g_{\tau\varpi}(\tilde{x}'\delta)}{t}\right)\right)\right]\\
&\leqslant_{(2)} 2|\widehat{\mathcal{T}}|\cdot|\widehat{\mathcal{W}}|\cdot|\widehat{\mathcal{N}}| \sup_{(\tau,\varpi,t)\in\widehat{\mathcal{T}}\times\widehat{\mathcal{W}}\times\widehat{\mathcal{N}}} \mathrm{E}\left[\exp\left(2\psi \sup_{\|\delta\|_{1,\varpi}=t} \mathbb{G}_n^o\left(\frac{K_\varpi(W)\tilde{x}'\delta}{t}\right)\right)\right]\\
&\leqslant_{(3)} 2|\widehat{\mathcal{T}}|\cdot|\widehat{\mathcal{W}}|\cdot|\widehat{\mathcal{N}}| \sup_{(\tau,\varpi,t)\in\widehat{\mathcal{T}}\times\widehat{\mathcal{W}}\times\widehat{\mathcal{N}}} \mathrm{E}\left[\exp\left(2\psi\left[\sup_{\|\delta\|_{1,\varpi}=t} \frac{\|\delta\|_{1,\varpi}}{t}\max_{j\leqslant p}\frac{|\mathbb{G}_n^o(K_\varpi(W)\tilde{x}_j)|}{\{\mathbb{E}_n[K_\varpi(W)\tilde{x}_j^2]\}^{1/2}}\right]\right)\right]\\
&=_{(4)} 2|\widehat{\mathcal{T}}|\cdot|\widehat{\mathcal{W}}|\cdot|\widehat{\mathcal{N}}| \sup_{(\tau,\varpi,t)\in\widehat{\mathcal{T}}\times\widehat{\mathcal{W}}\times\widehat{\mathcal{N}}} \mathrm{E}\left[\exp\left(2\psi\left[\max_{j\leqslant p}\frac{|\mathbb{G}_n^o(K_\varpi(W)\tilde{x}_j)|}{\{\mathbb{E}_n[K_\varpi(W)\tilde{x}_j^2]\}^{1/2}}\right]\right)\right]\\
&\leqslant_{(5)} 4p|\widehat{\mathcal{T}}|\cdot|\widehat{\mathcal{W}}|\cdot|\widehat{\mathcal{N}}| \max_{j\leqslant p}\sup_{\varpi\in\widehat{\mathcal{W}}} \mathrm{E}\left[\exp\left(4\psi\frac{\mathbb{G}_n^o(K_\varpi(W)\tilde{x}_j)}{\{\mathbb{E}_n[K_\varpi(W)\tilde{x}_j^2]\}^{1/2}}\right)\right]\\
&\leqslant_{(6)} 8p|\widehat{\mathcal{T}}|\cdot|\widehat{\mathcal{W}}|\cdot|\widehat{\mathcal{N}}|\exp\left(8\psi^2\right)
\end{aligned}$$

where (1) follows by $\exp(\max_{i\in I}|z_i|) \leqslant 2|I|\max_{i\in I}\exp(z_i)$, (2) by contraction principle (apply Theorem 4.12 [46] with $t_i = K_\varpi(W_i)\tilde{x}_i'\delta$, and $\phi_i(t_i) = \rho_\tau(K_\varpi(W_i)\tilde{y}_i - K_\varpi(W_i)\tilde{x}_i'\eta_\tau + t_i) - \rho_\tau(K_\varpi(W_i)\tilde{y}_i - K_\varpi(W_i)\tilde{x}_i'\eta_\tau)$ so that $|\phi_i(s)-\phi_i(t)| \leqslant |s-t|$ and $\phi_i(0)=0$ (3) follows by

$$|\mathbb{G}_n^o(K_\varpi(W_i)\tilde{x}_i'\delta)| \leqslant \|\delta\|_{1,\varpi}\max_{j\leqslant p}|\mathbb{G}_n^o(K_\varpi(W)\tilde{x}_j)/\{\mathbb{E}_n[K_\varpi(W)\tilde{x}_j^2]\}^{1/2}|,$$

(4) by definition of the suprema, (5) we again used $\exp(\max_{i\in I}|z_i|) \leqslant 2|I|\max_{i\in I}\exp(z_i)$, and (6) $\exp(z)+\exp(-z) \leqslant 2\exp(z^2/2)$.

■

Lemma 14 (Estimation Error of Refitted Quantile Regression)**.** *Consider an arbitrary vector $\widehat{\eta}_u$ and suppose $\|\eta_u\|_0 \leqslant s$. Let $\bar{r}_u \geqslant \|r_{ui}\|_{n,\varpi}$, $\widehat{s}_u \geqslant |\mathrm{support}(\widehat{\eta}_u)|$ and $\widehat{Q}_u \geqslant \mathbb{E}_n[K_\varpi(W)\{\rho_\tau(\tilde{y}_i - \tilde{x}_i'\widetilde{\eta}_u) - \rho_\tau(\tilde{y}_i - \tilde{x}_i'\eta_u)\}]$ for all $u\in\mathcal{U}$ hold. Furthermore, suppose that*

$$\sup_{u=(\tau,\varpi)\in\mathcal{U}}\left|\mathbb{E}_n\left(K_\varpi(W)\frac{\rho_\tau(\tilde{y}-\tilde{x}'\widetilde{\eta}_u)-\rho_\tau(\tilde{y}-\tilde{x}'\eta_u)}{\|\widetilde{\eta}_u-\eta_u\|_{1,\varpi}} - \mathrm{E}\left[K_\varpi(W)\frac{\rho_\tau(\tilde{y}-\tilde{x}'\widetilde{\eta}_u)-\rho_\tau(\tilde{y}-\tilde{x}'\eta_u)}{\|\widetilde{\eta}_u-\eta_u\|_{1,\varpi}} \mid W,\tilde{x}\right]\right)\right| \leqslant \frac{t_3}{\sqrt{n}}.$$

Then we have for n large enough,

$$\|\sqrt{f_u}\tilde{x}_i'(\widetilde{\eta}_u-\eta_u)\|_{n,\varpi} \lesssim \widetilde{N} := \sqrt{\frac{(\widehat{s}_u+s)}{\phi_{\min}(u,\widehat{s}_u+s)}}(K_{n1}+t_3/\sqrt{n}) + K_{n2} + \bar{f}\bar{r}_u + \widehat{Q}_u^{1/2}$$

where $\phi_{\min}(u,k) = \inf_{\|\delta\|_0=k}\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi}^2/\|\delta\|^2$, provided that

$$\sup_{u\in\mathcal{U},\|\bar{\delta}\|_0\leqslant\widehat{s}_u+s}\frac{\mathbb{E}_n[|r_{ui}||\tilde{x}_i'\bar{\delta}|^2]}{\mathbb{E}_n[|\tilde{x}_i'\bar{\delta}|^2]} + \widetilde{N}\sup_{\|\bar{\delta}\|_0\leqslant\widehat{s}_u+s}\frac{\mathbb{E}_n[|\tilde{x}_i'\bar{\delta}|^3]}{\mathbb{E}_n[|\tilde{x}_i'\bar{\delta}|^2]^{3/2}} \to 0.$$

Proof of Lemma 14. Let $\widehat{\delta}_u = \widetilde{\eta}_u - \eta_u$ which satisfies $\|\widehat{\delta}_u\|_0 \leqslant \widehat{s}_u + s$. By optimality of $\widetilde{\eta}_u$ in the refitted quantile regression we have with probability $1-\gamma$

$$\begin{aligned}
&\mathbb{E}_n[K_\varpi(W)\rho_\tau(\tilde{y}_i-\tilde{x}_i'\widetilde{\eta}_u)] - \mathbb{E}_n[K_\varpi(W)\rho_\tau(\tilde{y}_i-\tilde{x}_i'\eta_u)]\\
&\leqslant \mathbb{E}_n[K_\varpi(W)\rho_\tau(\tilde{y}_i-\tilde{x}_i'\widehat{\eta}_u)] - \mathbb{E}_n[K_\varpi(W)\rho_\tau(\tilde{y}_i-\tilde{x}_i'\eta_u)] \leqslant \widehat{Q}_u
\end{aligned} \tag{F.52}$$

Uniformly over $u\in\mathcal{U}$, we have conditional on $(W_i,\tilde{x}_i,r_{ui})_{i=1}^n$ that

$$\left|\mathbb{G}_n\left(K_\varpi(W)\frac{\rho_\tau(\tilde{y}-\tilde{x}'(\eta_u+\widetilde{\delta}_u))-\rho_\tau(\tilde{y}-\tilde{x}'\eta_u)}{\|\widetilde{\delta}_u\|_{1,\varpi}}\right)\right| \leqslant t_3. \tag{F.53}$$

Thus combining relations (F.52) and (F.53), we have

$$\mathbb{E}_n \mathrm{E}[K_\varpi(W)\{\rho_u(\tilde{y}_i - \tilde{x}_i'(\eta_u + \widetilde{\delta}_u)) - \rho_u(\tilde{y}_i - \tilde{x}_i'\eta_u)\} \mid \tilde{x}, \tilde{r}, \varpi] \quad \leqslant \quad \|\widetilde{\delta}_u\|_{1,\varpi} \frac{t_3}{\sqrt{n}} + \widehat{Q}_u$$

Invoking the sparse identifiability relation of Lemma 12, since $\sup_{\|\delta\|_0 \leqslant \widehat{s}_u + s} \frac{\mathbb{E}_n[|r_{ui}| \ |\tilde{x}_i'\theta|^2]}{\mathbb{E}_n[|\tilde{x}_i'\theta|^2]} \to 0$ by assumption, for n large enough

$$\frac{\|\sqrt{f_u}\tilde{x}'\widetilde{\delta}_u\|_{n,\varpi}^2}{4} \wedge \left\{\bar{q}_A \|\sqrt{f_u}\tilde{x}'\widetilde{\delta}_u\|_{n,\varpi}\right\} \quad \leqslant \quad K_{n2}\|\sqrt{f_u}\tilde{x}'\widetilde{\delta}_u\|_{n,\varpi} + \|\widetilde{\delta}_u\|_{1,\varpi}(K_{1n} + \frac{t_3}{\sqrt{n}}) + \widehat{Q}_u.$$

where $\bar{q}_A$ is defined with $A := \{\delta : \|\delta\|_0 \leqslant \widehat{s}_u + s\}$ and $\|\widetilde{\delta}_u\|_{1,\varpi} \leqslant \sqrt{(\widehat{s}_u + s)/\phi_{\min}(u, \widehat{s}_u + s_u)}\|\sqrt{f_u}\tilde{x}'\widetilde{\delta}_u\|_{n,\varpi}$.

Under the assumed growth condition, we have

$$\bar{q}_A < \widehat{Q}_u^{1/2} + K_{n2} + (K_{n1} + t_3/\sqrt{n})\sqrt{(\widehat{s}_u + s)/\phi_{\min}(u, \widehat{s}_u + s_u)}$$

and the minimum is achieved in the quadratic part. Therefore, for n sufficiently large, we have

$$\|\sqrt{f_u}\tilde{x}'\widetilde{\delta}_u\|_{n,\varpi} \leqslant \widehat{Q}_u^{1/2} + K_{n2} + (K_{n1} + t_3/\sqrt{n})\sqrt{(\widehat{s}_u + s)/\phi_{\min}(u, \widehat{s}_u + s_u)}.$$

■

Under the condition $\max_{i\leqslant n} \|\tilde{x}_i\|_\infty^2 \log(n \vee p) = o(n \min_{\tau\in\mathcal{T}} \tau(1-\tau))$, the next result provides new bounds for the data driven penalty choice parameter when the quantile indices in $\mathcal{T}$ can approach the extremes.

Lemma 15 (Pivotal Penalty Parameter Bound). *Let $\underline{\tau} = \min_{\tau\in\mathcal{T}} \tau(1-\tau)$ and $K_n = \max_{i\leqslant n, j\in[p]} |\tilde{x}_{ij}/\widehat{\sigma}_j|$, $\widehat{\sigma}_j = \mathbb{E}_n[\tilde{x}_j^2]^{1/2}$. Under $K_n^2 \log(p/\underline{\tau}) = o(n\underline{\tau})$, for n large enough we have that for some constant $\bar{C}$*

$$\Lambda(1-\alpha \mid \tilde{x}_1, \ldots, \tilde{x}_n) \leqslant \sqrt{1 + \frac{\log(16/\alpha)}{\log(d/\underline{\tau})}}\bar{C}\sqrt{\frac{\log(d/\underline{\tau})}{n}}$$

where $\Lambda(1-\alpha \mid \tilde{x}_1, \ldots, \tilde{x}_n)$ is the $1-\alpha$ quantile of $\sup_{\tau\in\mathcal{T}} \left|\frac{\sum_{i=1}^n \tilde{x}_j(\tau - 1\{U\leqslant\tau\})}{\widehat{\sigma}_j\sqrt{\tau(1-\tau)}}\right|$ conditional on $\tilde{x}_1, \ldots, \tilde{x}_n$, and U_i are independent uniform$(0,1)$ random variables.

Proof. Conditional on $\tilde{x}_1, \ldots, \tilde{x}_n$, letting $\widehat{\sigma}_j^2 = \mathbb{E}_n[x_j^2]$, we have that

$$n\Lambda = \sup_{\tau\in\mathcal{T}} \left|\frac{\sum_{i=1}^n \tilde{x}_j(\tau - 1\{U \leqslant \tau\})}{\widehat{\sigma}_j\sqrt{\tau(1-\tau)}}\right|.$$

Step 1. (Entropy Calculation) Let $\mathcal{F} = \{\tilde{x}_{ij}(\tau - 1\{U_i \leqslant \tau\})/\widehat{\sigma}_j : \tau \in \mathcal{T}, j \in [p]\}$, $h_\tau = \sqrt{\tau(1-\tau)}$, and $\mathcal{G} = \{f_\tau/h_\tau : \tau \in \mathcal{T}\}$. We have that

$$\begin{aligned} d(f_\tau/h_\tau, f_{\bar{\tau}}/h_{\bar{\tau}}) &\leqslant d(f_\tau, f_{\bar{\tau}})/h_\tau + d(f_{\bar{\tau}}/h_\tau, f_{\bar{\tau}}/h_{\bar{\tau}}) \\ &\leqslant d(f_\tau, f_{\bar{\tau}})/h_\tau + d(0, f_{\bar{\tau}}/h_{\bar{\tau}})|h_\tau - h_{\bar{\tau}}|/h_\tau \end{aligned}$$

Therefore, since $\|F\|_Q \leqslant \|G\|_Q$ by $h_\tau \leqslant 1$, and $d(0, f_{\bar{\tau}}/h_{\bar{\tau}}) \leqslant 1/h_{\bar{\tau}}$ we have

$$N(\varepsilon\|G\|_Q, \mathcal{G}, Q) \leqslant N(\varepsilon\|F\|_Q/\{2\min_{\tau\in\mathcal{T}} h_\tau\}, \mathcal{F}, Q)N(\varepsilon/\{2\min_{\tau\in\mathcal{T}} h_\tau^2\}, \mathcal{T}, |\cdot|).$$

Thus we have for some constants K and v that

$$N(\varepsilon\|G\|_Q,\mathcal{G},Q)\leqslant d(K/\{\epsilon\min_{\tau\in\mathcal{T}}h_\tau^2\})^v.$$

Step 2.(Symmetrization) Since we have $E[g^2]=1$ for all $g\in\mathcal{G}$, by Lemma 2.3.7 in [61] we have

$$P(\Lambda\geqslant t\sqrt{n})\quad\leqslant 4P(\max_{j\leqslant d}\sup_{\tau\in\mathcal{T}}|\mathbb{G}_n^o(g)|\geqslant t/4)$$

where $\mathbb{G}_n^o:\mathcal{G}\to\mathbb{R}$ is the symmetrized process generated by Rademacher variables. Conditional on $(x_1,u_1),\ldots,(x_n,u_n)$, we have that $\{\mathbb{G}_n^o(g):g\in\mathcal{G}\}$ is sub-Gaussian with respect to the $L_2(\mathbb{P}_n)$-norm by the Hoeffding inequality. Thus, by Lemma 16 in [10], for $\delta_n^2=\sup_{g\in\mathcal{G}}\mathbb{E}_n[g^2]$ and $\bar\delta_n=\delta_n/\|G\|_{\mathbb{P}_n}$, we have

$$P(\sup_{g\in\mathcal{G}}|\mathbb{G}_n^o(g)|>C\bar{K}\delta_n\sqrt{\log(dK/\underline{\mathcal{T}})}\mid\{\tilde{x}_i,U_i\}_{i=1}^n)\leqslant\int_0^{\bar\delta_n/2}\epsilon^{-1}\{d(K/\{\epsilon\min_{\tau\in\mathcal{T}}h_\tau^2\})^v\}^{-C^2+1}d\varepsilon$$

for some universal constant $\bar{K}$.

In order to control δ_n, note that $\delta_n^2=\sup_{g\in\mathcal{G}}\frac{1}{\sqrt{n}}\mathbb{G}_n(g^2)+E[g^2]$. In turn, since $\sup_{g\in\mathcal{G}}\mathbb{E}_n[g^4]\leqslant\delta_n^2\max_{i\leqslant n}G_i^2$, we have

$$P(\sup_{g\in\mathcal{G}}|\mathbb{G}_n^o(g^2)|>C\bar{K}\delta_n\max_{i\leqslant n}G_i\sqrt{\log(dK/\underline{\mathcal{T}})}\mid\{\tilde{x}_i,U_i\}_{i=1}^n)\leqslant\int_0^{\bar\delta_n/2}\epsilon^{-1}\{d(K/\{\epsilon\underline{\mathcal{T}}\})^v\}^{-C^2+1}d\varepsilon$$

Thus with probability $1-\int_0^{1/2}\epsilon^{-1}\{d(K/\epsilon\underline{\mathcal{T}})^v\}^{-C^2+1}d\varepsilon$, since $E[g^2]=1$ and $\max_{i\leqslant n}G_i\leqslant K_n/\sqrt{\underline{\mathcal{T}}}$, we have

$$\delta_n\leqslant 1+\frac{C'K_n\sqrt{\log(dK/\underline{\mathcal{T}})}}{\sqrt{n}\sqrt{\underline{\mathcal{T}}}}.$$

Therefore, under $K_n\sqrt{\log(dK/\underline{\mathcal{T}})}=o(\sqrt{n}\sqrt{\underline{\mathcal{T}}})$, conditionally on $\{\tilde{x}_i\}_{i=1}^n$ and n sufficiently large, with probability $1-2\int_0^{1/2}\epsilon^{-1}\{d(K/\{\epsilon\underline{\mathcal{T}}\})^v\}^{-C^2+1}d\varepsilon$ we have that

$$\sup_{g\in\mathcal{G}}|\mathbb{G}_n^o(g)|\leqslant 2C\bar{K}\sqrt{\log(dK/\underline{\mathcal{T}})}$$

The stated bound follows since for $C>2$

$$2\int_0^{1/2}\epsilon^{-1}\{d(K/\{\epsilon\underline{\mathcal{T}}\})^v\}^{-C^2+1}d\varepsilon\leqslant\{d/\underline{\mathcal{T}}\}^{-C^2+1}2\int_0^{1/2}\epsilon^{-2+C^2}d\varepsilon\leqslant\{d/\underline{\mathcal{T}}\}^{-C^2+1}.$$

■

Appendix G. Inequalities

Lemma 16. *Consider $\widehat\beta_u$ and β_u where $\|\beta_u\|_0\leqslant s$. Denote by $\widehat\beta^\lambda$ the vector with $\widehat\beta_j^\lambda=\widehat\beta_j1\{\widehat\sigma_{uj}|\widehat\beta_j|\geqslant\lambda\}$ where $\widehat\sigma_{uj}=\{\mathbb{E}_n[K_\varpi(W)(Z_j^a)^2]\}^{1/2}$. We have that*

$$\begin{aligned}\|\widehat\beta_u^\lambda-\beta_u\|_{1,\varpi}&\leqslant\|\widehat\beta_u-\beta_u\|_{1,\varpi}+s\lambda\\|\mathrm{support}(\widehat\beta_u^\lambda)|&\leqslant s+\|\widehat\beta_u-\beta_u\|_{1,\varpi}/\lambda\\\|Z^a(\widehat\beta_u^\lambda-\beta_u)\|_{n,\varpi}&\leqslant\|Z^a(\widehat\beta_u-\beta_u)\|_{n,\varpi}+\sqrt{\tilde\phi_{\max}(s,\varpi)}\{2\sqrt{s}\lambda+\|\widehat\beta_u-\beta_u\|_{1,\varpi}/\sqrt{s}\}\end{aligned}$$

where $\tilde{\phi}_{\max}(m,\varpi)=\sup_{1\leqslant\|\theta\|_0\leqslant m}\|\tilde{Z}^a\theta\|_{n,\varpi}/\|\theta\|$ *where* $\tilde{Z}^a_{ij}=Z^a_{ij}/\{\mathbb{E}_n[K_\varpi(W)(Z^a_j)^2]\}^{1/2}$.

Proof. Let $T_u=\text{support}(\beta_u)$. The first relation follows from the triangle inequality

$$\begin{aligned}\|\widehat{\beta}^\lambda_u-\beta_u\|_{1,\varpi} &= \|(\widehat{\beta}^\lambda_u-\beta_u)_{T_u}\|_{1,\varpi}+\|(\widehat{\beta}^\lambda_u)_{T_u^c}\|_{1,\varpi}\\ &\leqslant \|(\widehat{\beta}^\lambda_u-\widehat{\beta}_u)_{T_u}\|_{1,\varpi}+\|(\widehat{\beta}_u-\beta_u)_{T_u}\|_{1,\varpi}+\|(\widehat{\beta}^\lambda_u)_{T_u^c}\|_{1,\varpi}\\ &\leqslant \lambda s+\|(\widehat{\beta}_u-\beta_u)_{T_u}\|_{1,\varpi}+\|(\widehat{\beta}_u)_{T_u^c}\|_{1,\varpi}\\ &= \lambda s+\|\widehat{\beta}_u-\beta_u\|_{1,\varpi}\end{aligned}$$

To show the second result note that $\|\widehat{\beta}_u-\beta_u\|_{1,\varpi}\geqslant\{|\text{support}(\widehat{\beta}^\lambda_u)|-s\}\lambda$. Therefore,

$$|\text{support}(\widehat{\beta}^\lambda_u)|\leqslant s+\|\widehat{\beta}_u-\beta_u\|_{1,\varpi}/\lambda$$

which yields the result.

To show the third bound, we start using the triangle inequality

$$\|Z^a(\widehat{\beta}^\lambda_u-\beta_u)\|_{n,\varpi}\leqslant\|Z^a(\widehat{\beta}^\lambda_u-\widehat{\beta}_u)\|_{n,\varpi}+\|Z^a(\widehat{\beta}_u-\beta_u)\|_{n,\varpi}.$$

Without loss of generality assume that order the components is so that $|(\widehat{\beta}^\lambda_u-\widehat{\beta}_u)_j|\widehat{\sigma}_{uj}$ is decreasing. Let T_1 be the set of s indices corresponding to the largest values of $|(\widehat{\beta}^\lambda_u-\widehat{\beta}_u)_j|\widehat{\sigma}_{uj}$. Similarly define T_k as the set of s indices corresponding to the largest values of $|(\widehat{\beta}^\lambda_u-\widehat{\beta}_u)_j|\widehat{\sigma}_{uj}$ outside $\cup_{m=1}^{k-1}T_m$. Therefore, $\widehat{\beta}^\lambda_u-\widehat{\beta}_u=\sum_{k=1}^{\lceil p/s\rceil}(\widehat{\beta}^\lambda_u-\widehat{\beta}_u)_{T_k}$. Moreover, given the monotonicity of the components, $\|(\widehat{\beta}^\lambda_u-\widehat{\beta}_u)_{T_k}\|_{2,\varpi}\leqslant\|(\widehat{\beta}^\lambda_u-\widehat{\beta}_u)_{T_{k-1}}\|_1/\sqrt{s}$. Then, we have

$$\begin{aligned}\|Z^a(\widehat{\beta}^\lambda_u-\widehat{\beta}_u)\|_{n,\varpi} &= \|Z^a\textstyle\sum_{k=1}^{\lceil p/s\rceil}(\widehat{\beta}^\lambda_u-\widehat{\beta}_u)_{T_k}\|_{n,\varpi}\\ &\leqslant \|Z^a(\widehat{\beta}^\lambda_u-\widehat{\beta}_u)_{T_1}\|_{n,\varpi}+\textstyle\sum_{k\geqslant2}\|Z^a(\widehat{\beta}^\lambda_u-\widehat{\beta}_u)_{T_k}\|_{n,\varpi}\\ &\leqslant \sqrt{\tilde{\phi}_{\max}(s,\varpi)}\|(\widehat{\beta}^\lambda_u-\widehat{\beta}_u)_{T_1}\|_{2,\varpi}+\sqrt{\tilde{\phi}_{\max}(s,\varpi)}\textstyle\sum_{k\geqslant2}\|(\widehat{\beta}^\lambda_u-\widehat{\beta}_u)_{T_k}\|_{2,\varpi}\\ &\leqslant \sqrt{\tilde{\phi}_{\max}(s,\varpi)}\lambda\sqrt{s}+\sqrt{\tilde{\phi}_{\max}(s,\varpi)}\textstyle\sum_{k\geqslant1}\|(\widehat{\beta}^\lambda_u-\widehat{\beta}_u)_{T_k}\|_{1,\varpi}/\sqrt{s}\\ &= \sqrt{\tilde{\phi}_{\max}(s,\varpi)}\lambda\sqrt{s}+\sqrt{\tilde{\phi}_{\max}(s,\varpi)}\|\widehat{\beta}^\lambda_u-\widehat{\beta}_u\|_{1,\varpi}/\sqrt{s}\\ &\leqslant \sqrt{\tilde{\phi}_{\max}(s,\varpi)}\{2\sqrt{s}\lambda+\|\widehat{\beta}_u-\beta_u\|_{1,\varpi}/\sqrt{s}\}\end{aligned}$$

where the last inequality follows from the first result and the triangle inequality.

∎

Lemma 17 (Supremum of Sparse Vectors on Symmetrized Random Matrices)**.** *Let* $\widehat{\mathcal{U}}$ *denote a finite set and* $(X_{ui})_{u\in\widehat{\mathcal{U}}}$, $i=1,\dots,n$, *be fixed vectors such that* $X_{ui}\in\mathbb{R}^p$ *and* $\max_{1\leqslant i\leqslant n}\max_{u\in\widehat{\mathcal{U}}}\|X_{ui}\|_\infty\leqslant K$. *Furthermore define*

$$\delta_n:=\bar{C}K\sqrt{k}\left(\sqrt{\log|\widehat{\mathcal{U}}|}+\sqrt{1+\log p}+\log k\sqrt{\log(p\vee n)}\sqrt{\log n}\right)/\sqrt{n},$$

where $\bar{C}$ *is a universal constant. Then,*

$$\mathrm{E}\left[\sup_{\|\theta\|_0\leqslant k,\|\theta\|=1}\max_{u\in\widehat{\mathcal{U}}}\left|\mathbb{E}_n\left[\epsilon(\theta'X_u)^2\right]\right|\right]\leqslant\delta_n\sup_{\|\theta\|_0\leqslant k,\|\theta\|=1,u\in\widehat{\mathcal{U}}}\sqrt{\mathbb{E}_n[(\theta'X_u)^2]}.$$

Proof. See [12] for the proof. ∎

Corollary 2 (Supremum of Sparse Vectors on Many Random Matrices)**.** *Let $\widehat{\mathcal{U}}$ denote a finite set and $(X_{ui})_{u\in\widehat{\mathcal{U}}}$, $i=1,\ldots,n$, be independent (across i) random vectors such that $X_{ui}\in\mathbb{R}^p$ and*

$$\sqrt{\mathrm{E}[\max_{1\leqslant i\leqslant n}\max_{u\in\widehat{\mathcal{U}}}\|X_{ui}\|_\infty^2]}\leqslant K.$$

Furthermore define

$$\delta_n := \bar{C}K\sqrt{k}\left(\sqrt{\log|\widehat{\mathcal{U}}|}+\sqrt{1+\log p}+\log k\sqrt{\log(p\vee n)}\sqrt{\log n}\right)/\sqrt{n},$$

where $\bar{C}$ is a universal constant. Then,

$$\mathrm{E}\left[\sup_{\|\theta\|_0\leqslant k,\|\theta\|=1}\max_{u\in\widehat{\mathcal{U}}}\left|\mathbb{E}_n\left[(\theta' X_u)^2-\mathrm{E}[(\theta' X_u)^2]\right]\right|\right]\leqslant \delta_n^2+\delta_n\sup_{\|\theta\|_0\leqslant k,\|\theta\|=1,u\in\widehat{\mathcal{U}}}\sqrt{\mathbb{E}_n\mathrm{E}[(\theta' X_u)^2]}.$$

We will also use the following result of [25].

Lemma 18 (Maximal Inequality)**.** *Work with the setup above. Suppose that $F\geqslant\sup_{f\in\mathcal{F}}|f|$ is a measurable envelope for $\mathcal{F}$ with $\|F\|_{P,q}<\infty$ for some $q\geqslant 2$. Let $M=\max_{i\leqslant n}F(W_i)$ and $\sigma^2>0$ be any positive constant such that $\sup_{f\in\mathcal{F}}\|f\|_{P,2}^2\leqslant\sigma^2\leqslant\|F\|_{P,2}^2$. Suppose that there exist constants $a\geqslant e$ and $v\geqslant 1$ such that*

$$\log\sup_Q N(\epsilon\|F\|_{Q,2},\mathcal{F},\|\cdot\|_{Q,2})\leqslant v\log(a/\epsilon),\ 0<\epsilon\leqslant 1.$$

Then

$$\mathrm{E}_P[\sup_{f\in\mathcal{F}}|\mathbb{G}_n(f)|]\leqslant K\left(\sqrt{v\sigma^2\log\left(\frac{a\|F\|_{P,2}}{\sigma}\right)}+\frac{v\|M\|_{P,2}}{\sqrt{n}}\log\left(\frac{a\|F\|_{P,2}}{\sigma}\right)\right),$$

where K is an absolute constant. Moreover, for every $t\geqslant 1$, with probability $>1-t^{-q/2}$,

$$\sup_{f\in\mathcal{F}}|\mathbb{G}_n(f)| \leqslant (1+\alpha)\mathrm{E}_P[\sup_{f\in\mathcal{F}}|\mathbb{G}_n(f)|]+K(q)\Big[(\sigma+n^{-1/2}\|M\|_{P,q})\sqrt{t}+\alpha^{-1}n^{-1/2}\|M\|_{P,2}t\Big],$$

$\forall\alpha>0$ where $K(q)>0$ is a constant depending only on q. In particular, setting $a\geqslant n$ and $t=\log n$, with probability $>1-c(\log n)^{-1}$,

$$\sup_{f\in\mathcal{F}}|\mathbb{G}_n(f)|\leqslant K(q,c)\left(\sigma\sqrt{v\log\left(\frac{a\|F\|_{P,2}}{\sigma}\right)}+\frac{v\|M\|_{P,q}}{\sqrt{n}}\log\left(\frac{a\|F\|_{P,2}}{\sigma}\right)\right), \tag{G.54}$$

where $\|M\|_{P,q}\leqslant n^{1/q}\|F\|_{P,q}$ and $K(q,c)>0$ is a constant depending only on q and c.

Appendix H. Confidence Regions for Function-Valued Parameters Based on Moment Conditions

For completeness, in this section we collect (simple adaptation of) the results of [12] that are invoked in our proofs. We are interested in function-valued target parameters indexed by $u\in\mathcal{U}\subset\mathbb{R}^{d_u}$. The true value of the target parameter is denoted by

$$\theta^0=(\theta_{uj})_{u\in\mathcal{U},j\in[\tilde{p}]},\quad\text{where}\ \ \theta_{uj}\in\Theta_{uj}\ \text{for each}\ u\in\mathcal{U}\ \ \text{and}\ \ j\in[\tilde{p}].$$

For each $u \in \mathcal{U}$ and $j \in [\tilde{p}]$, the parameter θ_{uj} is characterized as the solution to the following moment condition:

$$\mathrm{E}[\psi_{uj}(W_{uj}, \theta_{uj}, \eta_{uj})] = 0, \tag{H.55}$$

where W_{uj} is a random vector that takes values in a Borel set $\mathcal{W}_{uj} \subset \mathbb{R}^{d_w}$, $\eta^0 = (\eta_{uj})_{u\in\mathcal{U},j\in[\tilde{p}]}$ is a nuisance parameter where $\eta_{uj} \in T_{uj}$ a convex set, and the moment function

$$\psi_{uj} : \mathcal{W}_{uj} \times \Theta_{uj} \times T_{ujn} \mapsto \mathbb{R}, \quad (w, \theta, t) \mapsto \psi_{uj}(w, \theta, t) \tag{H.56}$$

is a Borel measurable map.

We assume that the (continuum) nuisance parameter η^0 can be modelled and estimated by $\widehat{\eta} = (\widehat{\eta}_{uj})_{u\in\mathcal{U},j\in[\tilde{p}]}$. We will discuss examples where the corresponding η^0 can be estimated using modern regularization and post-selection methods such as Lasso and Post-Lasso (although other procedures can be applied). The estimator $\check{\theta}_{uj}$ of θ_{uj} is constructed as any approximate ϵ_n-solution in Θ_{uj} to a sample analog of the moment condition (H.55), i.e.,

$$\max_{j\in[\tilde{p}]} \sup_{u\in\mathcal{U}} \left\{ |\mathbb{E}_n[\psi_{uj}(W_{uj}, \check{\theta}_{uj}, \widehat{\eta}_{uj})]| - \inf_{\theta_j\in\Theta_{uj}} |\mathbb{E}_n[\psi_{uj}(W_{uj}, \theta, \widehat{\eta}_{uj})]| \right\} \leqslant \epsilon_n = o_P(n^{-1/2}\delta_n). \tag{H.57}$$

As discussed before, we rely on an orthogonality condition for regular estimation of θ_{uj} that we state next.

Definition 1 (**Near Orthogonality condition**)**.** For each $u \in \mathcal{U}$ and $j \in [\tilde{p}]$, we say that ψ_{uj} obeys a general form of orthogonality with respect to $\mathcal{H}_{uj}$ uniformly in $u \in \mathcal{U}$, if the following conditions hold: The Gateaux derivative map

$$\mathrm{D}_{u,j,\bar{r}}[\tilde{\eta}_{uj} - \eta_{uj}] := \partial_r \mathrm{E}\left(\psi_{uj}\left\{ W_{uj}, \theta_{uj}, \eta_{uj} + r\left[\tilde{\eta}_{uj} - \eta_{uj}\right] \right\} \right)\Bigg|_{r=\bar{r}}$$

exists for all $r \in [0, 1)$, $\tilde{\eta} \in \mathcal{H}_{uj}$, $j \in \tilde{p}$, and $u \in \mathcal{U}$ and vanishes at $r = 0$, namely,

$$|\mathrm{D}_{u,j,0}[\tilde{\eta}_{uj} - \eta_{uj}]| \leqslant \delta_n n^{-1/2} \quad \text{for all } \tilde{\eta}_{uj} \in \mathcal{H}_{uj}. \tag{H.58}$$

In what follows, we shall denote by c_0, c, and C some positive constants.

Assumption C 1 (Moment condition problem)**.** *Consider a random element W, taking values in a measure space $(\mathcal{W}, \mathcal{A}_{\mathcal{W}})$, with law determined by a probability measure $P \in \mathcal{P}_n$. The observed data $((W_{ui})_{u\in\mathcal{U}})_{i=1}^n$ consist of n i.i.d. copies of a random element $(W_u)_{u\in\mathcal{U}}$ which is generated as a suitably measurable transformation with respect to W and u. Uniformly for all $n \geqslant n_0$ and $P \in \mathcal{P}_n$, the following conditions hold: (i) The true parameter value θ_{uj} obeys (H.55) and is interior relative to Θ_{uj}, namely there is a ball of radius $Cn^{-1/2}u_n \log n$ centered at θ_{uj} contained in Θ_{uj} for all $u \in \mathcal{U}$, $j \in [\tilde{p}]$ where $u_n := \mathrm{E}[\sup_{u\in\mathcal{U},j\in[\tilde{p}]} |\sqrt{n}\mathbb{E}_n[\psi_{uj}(W_{uj}, \theta_{uj}, \eta_{uj})]|]$. (ii) For each $u \in \mathcal{U}$ and $j \in [\tilde{p}]$, the map $(\theta, \eta) \in \Theta_{uj} \times \mathcal{H}_{uj} \mapsto \mathrm{E}[\psi_{uj}(W_{uj}, \theta, \eta)]|$ is twice continuously differentiable. (iii) For all $u \in \mathcal{U}$ and $j \in [\tilde{p}]$, the moment function ψ_{uj} obeys the orthogonality condition given in Definition 1 for the set $\mathcal{H}_{uj} = \mathcal{H}_{ujn}$ specified in Assumption 2. (iv) The following identifiability condition holds: $|\mathrm{E}[\psi_{uj}(W_{uj}, \theta, \eta_{uj})]| \geqslant \frac{1}{2}|J_{uj}(\theta - \theta_{uj})| \wedge c_0$ for all $\theta \in \Theta_{uj}$, where $J_{uj} := \partial_\theta \mathrm{E}[\psi_{uj}(W_{uj}, \theta, \eta_{uj})]|_{\theta=\theta_{uj}}$ satisfies $0 < j_n < |J_{uj}| < C < \infty$ for all $u \in \mathcal{U}$ and $j \in [\tilde{p}]$. (v) The following smoothness conditions holds*

(a) $\sup_{u\in\mathcal{U},j\in[\tilde p],(\theta,\bar\theta)\in\Theta_{uj}^2,(\eta,\bar\eta)\in\mathcal{H}_{ujn}^2} \frac{\mathrm{E}[\{\psi_{uj}(W_{uj},\theta,\eta)-\psi_{uj}(W_{uj},\bar\theta,\bar\eta)\}^2]}{\{|\theta-\bar\theta|\vee\|\eta-\bar\eta\|_e\}^\alpha} \leqslant C,$

(b) $\sup_{u\in\mathcal{U},(\theta,\eta)\in\Theta_{uj}\times\mathcal{H}_{ujn},r\in[0,1)} |\partial_r \mathrm{E}\left[\psi_{uj}(W_{uj},\theta,\eta_{uj}+r\{\eta-\eta_{uj}\})\right]|/\|\eta-\eta_{uj}\|_e \leqslant \bar B_{1n},$

(c) $\sup_{u\in\mathcal{U},j\in[\tilde p],(\theta,\eta)\in\Theta_{uj}\times\mathcal{H}_{ujn},r\in[0,1)} \frac{|\partial_r^2\mathrm{E}[\psi_{uj}(W_{uj},\theta_{uj}+r\{\theta-\theta_{uj}\},\eta_{uj}+r\{\eta-\eta_{uj}\})]|}{\{|\theta-\theta_{uj}|^2\vee\|\eta-\eta_{uj}\|_e^2\}} \leqslant \bar B_{2n}.$

Next we state assumptions on the nuisance functions. In what follows, let $\Delta_n \searrow 0$, $\delta_n \searrow 0$, and $\tau_n \searrow 0$ be sequences of constants approaching zero from above at a speed at most polynomial in n (for example, $\delta_n \geqslant 1/n^c$ for some $c>0$).

Assumption C2 (Estimation of nuisance functions)**.** *The following conditions hold for each $n \geqslant n_0$ and all $P\in\mathcal{P}_n$. The estimated functions $\widehat\eta_{uj}\in\mathcal{H}_{ujn}$ with probability at least $1-\Delta_n$, where $\mathcal{H}_{ujn}$ is the set of measurable maps $\tilde\eta_{uj}$ such that*

$$\sup_{u\in\mathcal{U}}\max_{j\in[\tilde p]}\|\tilde\eta_{uj}-\eta_{uj}\|_e \leqslant \tau_n,$$

where the e-norm is the same in Assumption 1, and whose complexity does not grow too quickly in the sense that $\mathcal{F}_1=\{\psi_{uj}(W_{uj},\theta,\eta): j\in[\tilde p], u\in\mathcal{U}, \theta\in\Theta_{uj}, \eta\in\mathcal{H}_{ujn}\cup\{\eta_{uj}\}\}$ is suitably measurable and its uniform covering entropy obeys:

$$\sup_Q \log N(\epsilon\|F_1\|_{Q,2},\mathcal{F}_1,\|\cdot\|_{Q,2}) \leqslant s_{n(\mathcal{U},\tilde p)}(\log(a_n/\epsilon))\vee 0,$$

where $F_1(W)$ is an envelope for $\mathcal{F}_1$ which is measurable with respect to W and satisfies $F_1(W)\geqslant \sup_{u\in\mathcal{U},j\in[\tilde p],\eta\in\mathcal{H}_{ujn},\theta\in\Theta_{uj}}|\psi_{uj}(W_{uj},\theta,\eta)|$ and $\|F_1\|_{P,q}\leqslant K_n$ for $q\geqslant 2$. The complexity characteristics $a_n\geqslant\max(n,K_n,\mathrm{e})$ and $s_{n(\mathcal{U},\tilde p)}\geqslant 1$ obey the growth conditions:

$$\begin{aligned} n^{-1/2}\sqrt{s_{n(\mathcal{U},\tilde p)}\log(a_n)} + n^{-1}s_{n(\mathcal{U},\tilde p)}n^{\frac1q}K_n\log(a_n) &\leqslant \tau_n\\ \{(1\vee\bar B_{1n})(\tau_n/j_n)\}^{\alpha/2}\sqrt{s_{n(\mathcal{U},\tilde p)}\log(a_n)} + s_{n(\mathcal{U},\tilde p)}n^{\frac1q-\frac12}K_n\log(a_n)\log n &\leqslant \delta_n,\\ \text{and}\quad \sqrt{n}\bar B_{2n}(1\vee\bar B_{1n})(\tau_n/j_n)^2 &\leqslant \delta_n \end{aligned}$$

where $\bar B_{1n}$, $\bar B_{2n}$, j_n, q and α are defined in Assumption 1.

Theorem 5 (Uniform Bahadur representation for a Continuum of Target Parameters)**.** *Under Assumptions 1 and 2, for an estimator $(\check\theta_{uj})_{u\in\mathcal{U},j\in[\tilde p]}$ that obeys equation (H.57),*

$$\sqrt{n}\sigma_{uj}^{-1}(\check\theta_{uj}-\theta_{uj}) = \mathbb{G}_n\bar\psi_{uj} + O_P(\delta_n) \text{ in } \ell^\infty(\mathcal{U}\times[\tilde p]), \text{ uniformly in } P\in\mathcal{P}_n,$$

where $\bar\psi_{uj}(W) := -\sigma_{uj}^{-1}J_{uj}^{-1}\psi_{uj}(W_{uj},\theta_{uj},\eta_{uj})$ and $\sigma_{uj}^2=\mathrm{E}[J_{uj}^{-2}\psi_{uj}^2(W_{uj},\theta_{uj},\eta_{uj})]$.

The uniform Bahadur representation derived in Theorem 5 is useful in the construction of simultaneous confidence bands for $(\theta_{uj})_{u\in\mathcal{U},j\in[\tilde p]}$. This is achieved by new high-dimensional central limit theorems that have been recently developed in [24] and [25]. We will make use of the following regularity condition. In what follows $\bar\delta_n$ and Δ_n are fixed sequences going to zero, and we denote by $\widehat\psi_{uj}(W_i) := -\widehat\sigma_{uj}^{-1}\widehat J_{uj}^{-1}\psi_{uj}(W_{uji},\check\theta_{uj},\widehat\eta_{uj})$ estimators of $\bar\psi_{uj}(W)$, with $\widehat J_{uj}$ and $\widehat\sigma_{uj}$ being suitable estimators of J_{uj} and σ_{uj}. In what follows, $\|\cdot\|_{\mathbb{P}_n,2}$ denotes the empirical $L_2(\mathbb{P}_n)$-norm where $\mathbb{P}_n$ is the empirical measure of the data.

Assumption C3 (Score Regularity)**.** *The following conditions hold for each $n \geqslant n_0$ and all $P \in \mathcal{P}_n$. (i) The class of function induced by the score $\mathcal{F}_0 = \{\bar\psi_{uj}(W) : j \in [\tilde p], u \in \mathcal{U}\}$ is suitably measurable and its uniform covering entropy obeys:*

$$\sup_Q \log N(\epsilon \|F_0\|_{Q,2}, \mathcal{F}_0, \|\cdot\|_{Q,2}) \leqslant \varrho_n(\log(A_n/\epsilon)) \vee 0,$$

where $F_0(W)$ is an envelope for $\mathcal{F}_0$ which is measurable with respect to W and satisfies $F_0(W) \geqslant \sup_{u\in\mathcal{U},j\in[\tilde p]} |\bar\psi_{uj}(W)|$ and $\|F_0\|_{P,q} \leqslant L_n$ for $q \geqslant 4$. Furthermore, $c \leqslant \sup_{u\in\mathcal{U},j\in[\tilde p]} \mathrm{E}[|\bar\psi_{uj}(W)|^k] \leqslant CL_n^{k-2}$ for $k = 2,3,4$. (ii) The set $\widehat{\mathcal{F}}_0 = \{\bar\psi_{uj}(W) - \widehat\psi_{uj}(W) : j \in [\tilde p], u \in \mathcal{U}\}$ satisfies the conditions $\log N(\epsilon, \widehat{\mathcal{F}}_0, \|\cdot\|_{\mathbb{P}_n,2}) \leqslant \bar\varrho_n(\log(\bar A_n/\epsilon)) \vee 0$, and $\sup_{u\in\mathcal{U},j\in[\tilde p]} \mathbb{E}_n[\{\bar\psi_{uj}(W) - \widehat\psi_{uj}(W)\}^2] \leqslant \bar\delta_n\{\rho_n\bar\rho_n \log(A_n \vee n)\log(\bar A_n \vee n)\}^{-1}$ with probability $1-\Delta_n$.

Assumption 3 imposes condition on the class of functions induces by $\bar\psi_{uj}$ and on its estimators $\widehat\psi_{uj}$. Typically the bound L_n on the moment of the envelope is smaller than K_n, and in many settings $\bar\rho_n = \rho_n \lesssim d_{\mathcal{U}}$ the dimension of $\mathcal{U}$.

Next let $\mathcal{N}$ denote a mean zero Gaussian process indexed by $\mathcal{U} \times [\tilde p]$ with covariance operator given by $\mathrm{E}[\bar\psi_{uj}(W)\bar\psi_{u'j'}(W)]$ for $j, j' \in [\tilde p]$ and $u, u' \in \mathcal{U}$. Because of the high-dimensionality, indeed $\tilde p$ can be larger than the sample size n, the central limit theorem will be uniformly valid over "rectangles." This class of sets are rich enough to construct many confidence regions of interest in applications accounting for multiple testing. Let $\mathcal{R}$ denote the set of rectangles $R = \{z \in \mathbb{R}^{\tilde p} : \max_{j\in A} z_j \leqslant t, \max_{j\in B}(-z_j) \leqslant t\}$ for all $A, B \subset [\tilde p]$ and $t \in \mathbb{R}$. The following result is a consequence of Theorem 3 above and Corollary 2.2 of [26].

Corollary 3. *Under Assumptions 1 and 2 with $\delta_n = o(\{\rho_n \log(A_n \vee n)\}^{-1/2})$, Assumption 3(i), and $\rho_n \log(A_n \vee n) = o(\{(n/L_n^2)^{1/7} \wedge (n^{1-2/q}/L_n^2)^{1/3}\})$, we have that*

$$\sup_{P\in\mathcal{P}_n}\sup_{R\in\mathcal{R}} \left| \mathrm{P}\left(\{\sup_{u\in\mathcal{U}} n^{1/2}\sigma_{uj}^{-1}(\check\theta_{uj} - \theta_{uj})\}_{j=1}^{\tilde p} \in R \right) - \mathrm{P}(\mathcal{N} \in R) \right| = o(1).$$

In order to derive a method to build confidence regions we approximate the process $\mathcal{N}$ by the Gaussian multiplier bootstrap based on estimates $\widehat\psi_{uj}$ of $\bar\psi_{uj}$, namely

$$\widehat{\mathcal{G}} = (\widehat{\mathcal{G}}_{uj})_{u\in\mathcal{U},j\in[\tilde p]} = \left\{ \frac{1}{\sqrt n} \sum_{i=1}^n \xi_i \widehat\psi_{uj}(W_i) \right\}_{u\in\mathcal{U},j\in[\tilde p]}$$

where $(\xi_i)_{i=1}^n$ are independent standard normal random variables which are independent from the data $(W_i)_{i=1}^n$. Based on Theorem 5.2 of [24], the following result shows that the multiplier bootstrap provides a valid approximation to the large sample probability law of $\sqrt n(\check\theta_{uj} - \theta_{uj})_{u\in\mathcal{U},j\in[\tilde p]}$ over rectangles.

Corollary 4 (**Uniform Validity of Gaussian Multiplier Bootstrap**)**.** *Under Assumptions 1 and 2 with $\delta_n = o(\{\rho_n \log(A_n \vee n)\}^{-1/2})$, Assumption 3, and $\rho_n \log(A_n \vee n) = o(\{(n/L_n^2)^{1/7} \wedge (n^{1-2/q}/L_n^2)^{1/3}\})$, we have that*

$$\sup_{P\in\mathcal{P}_n}\sup_{R\in\mathcal{R}} \left| \mathrm{P}\left(\{\sup_{u\in\mathcal{U}} n^{1/2}\sigma_{uj}^{-1}(\check\theta_{uj} - \theta_{uj})\}_{j=1}^{\tilde p} \in R \right) - \mathrm{P}(\widehat{\mathcal{G}} \in R \mid (W_i)_{i=1}^n) \right| = o(1)$$

Appendix I. Continuum of ℓ_1-Penalized M-Estimators

For the reader's convenience, this section collects results on the estimation of a continuum of estimation of high-dimensional models via ℓ_1-penalized estimators.

Consider a data generating process with a response variable $(Y_u)_{u\in\mathcal{U}}$ and observable covariates $(X_u)_{u\in\mathcal{U}}$ satisfying for each $u \in \mathcal{U}$,

$$\theta_u \in \arg\min_{\theta\in\mathbb{R}^p} \mathrm{E}[M_u(Y_u, X_u, \theta, a_u)], \tag{I.59}$$

where θ_u is a p-dimensional vector, a_u is a nuisance function that capture the misspecification of the model, M_u is a pre-specified function, and the (p_u-dimensional, $p_u \leqslant p$) covariate X_u could have been constructed based on transformations of other variables. This implies that

$$\partial_\theta \mathrm{E}[M_u(Y_u, X_u, \theta_u, a_u)] = 0 \text{ for all } u \in \mathcal{U}.$$

The solution θ_u is assumed to be sparse in the sense that for some process $(\theta_u)_{u\in\mathcal{U}}$ satisfies

$$\|\theta_u\|_0 \leqslant s \text{ for all } u \in \mathcal{U}.$$

Because the nuisance function, such sparsity assumption is very mild and formulation (I.59) encompasses several cases of interest including approximate sparse models. We focus on the estimation of $(\theta_u)_{u\in\mathcal{U}}$ and we assume that an estimate $\widehat{a}_u$ of the nuisance function a_u is available and the criterion $M_u(Y_u, X_u, \theta_u) := M_u(Y_u, X_u, \theta_u, \widehat{a}_u)$ is used as a proxy for $M_u(Y_u, X, \theta_u, a_u)$.

In the case of linear regression we have $M_u(y, x, \theta) = \frac{1}{2}(y - x'\theta)^2$. In the logistic regression case, we have $M_u(y, x, \theta) = -\{1(y = 1) \log G(x'\theta) + 1(y = 0) \log(1 - G(x'\theta))\}$ where G is the logistic link function $G(t) = \exp(t)/\{1 + \exp(t)\}$. Additional examples include quantile regression models where for $u \in (0, 1)$ we have .

Example 8 (Quantile Regression Model)**.** Consider a data generating process $Y = F^{-1}_{Y|X}(U) = X'\theta_U + r_U(X)$, where $U \sim \mathrm{Unif}(0, 1)$, and X is a p-dimensional vector of covariates. The criterion $M_u(y, x, \theta) = (u - 1\{y \leqslant x'\theta\})(y - x'\theta)$ with the (trivial) estimate $\widehat{a}_u = 0$ of the nuisance parameter $a_u = r_u$.

Example 9 (Lasso with Estimated Weights)**.** We consider a linear model defined as $f_uY = f_uX'\theta_u + \bar{r}_u + \zeta_u, \quad \mathrm{E}[f_uX\zeta_u] = 0$, where X are $\bar{p}$-dimensional covariates, θ_u is a s-sparse vector, and $\bar{r}_u$ is an approximation error satisfying $\sup_{u\in\mathcal{U}} \mathbb{E}_n[\bar{r}_u^2] \lesssim_P s \log \bar{p}/n$. In this setting, (Y, X) are observed and only an estimator $\widehat{f}_u$ of f_u is available. This corresponds to a nuisance parameter $a_u = (f_u, \bar{r}_u)$ and $\widehat{a}_u = (\widehat{f}_u, 0)$ so that $\mathbb{E}_n[M_u(Y, X, \theta, a_u)] = \mathbb{E}_n[f_u^2(Y - X'\theta - \bar{r}_u)^2]$ and $\mathbb{E}_n[M_u(Y, X, \theta)] = \mathbb{E}_n[\widehat{f}_u^2(Y - X'\theta)^2]$.

We assume that n i.i.d. observations from dgps where (I.59) holds, $\{(Y_{ui}, X_{ui})_{u\in\mathcal{U}}\}_{i=1}^n$, are available to estimate $(\theta_u)_{u\in\mathcal{U}}$. For each $u \in \mathcal{U}$, a penalty level λ, and a diagonal matrix of penalty loadings $\widehat{\Psi}_u$, we define the ℓ_1-penalized M_u-estimator (Lasso) as

$$\widehat{\theta}_u \in \arg\min_\theta \mathbb{E}_n[M_u(Y_u, X_u, \theta)] + \frac{\lambda}{n}\|\widehat{\Psi}_u\theta\|_1. \tag{I.60}$$

Furthermore, for each $u \in \mathcal{U}$, the post-penalized estimator (Post-Lasso) based on a set of covariates $\widetilde{T}_u$ is then defined as

$$\widetilde{\theta}_u \in \arg\min_\theta \mathbb{E}_n[M_u(Y_u, X_u, \theta)] \quad : \quad \mathrm{support}(\theta) \subseteq \widetilde{T}_u. \tag{I.61}$$

Potentially, the set $\widetilde{T}_u$ contains support$(\widehat{\theta}_u)$ and possibly additional variables deemed as important (although in that case the total number of additional variables should also obey the same growth conditions that s obeys). We will set $\widetilde{T}_u = \text{support}(\widehat{\theta}_u)$ unless otherwise noted.

In order to handle the functional response data, the penalty level λ and penalty loading $\widehat{\Psi}_u = \text{diag}(\{\widehat{l}_{uk}, k = 1, \ldots, p\})$ need to be set to control selection errors uniformly over $u \in \mathcal{U}$. The choice of loading matrix is problem specific and we suggest to mimic the following "ideal" choice $\widehat{\Psi}_{u0} = \text{diag}(\{l_{uk}, k = 1, \ldots, p\})$ where

$$l_{uk} = \{\mathbb{E}_n\left[\{\partial_{\theta_k} M_u(Y_u, X_u, \theta_u, a_u)\}^2\right]\}^{1/2} \tag{I.62}$$

which is motivated by the use of self-normalized moderate deviation theory. In that case it is suitable to set λ so that with high probability

$$\frac{\lambda}{n} \geqslant c \sup_{u\in\mathcal{U}} \left\| \widehat{\Psi}_{u0}^{-1} \mathbb{E}_n\left[\partial_\theta M_u(Y_u, X_u, \theta_u, a_u)\right] \right\|_\infty, \tag{I.63}$$

where $c > 1$ is a fixed constant. Indeed, in the case that $\mathcal{U}$ is a singleton the choice above is similar to [22], [11], and [18]. This approach was first employed for a continuum of indices $\mathcal{U}$ in the context of ℓ_1-penalized quantile regression processes by [10].

To implement (I.63), we propose setting the penalty level as

$$\lambda = c\sqrt{n}\Phi^{-1}(1 - \gamma/\{2pN_n\}), \tag{I.64}$$

where N_n is a measure of the class of functions indexed by $\mathcal{U}$, $1 - \gamma$ (with $\gamma = o(1)$) is a confidence level associated with the probability of event (I.63), and $c > 1$ is a slack constant. In many settings we can take $N_n = n^{d_\mathcal{U}}$. If the set $\mathcal{U}$ is a singleton, $N_n = 1$ suffices which corresponds to that used in [15]. When implementing the estimators, we set $c = 1.1$ and $\gamma = .1/\log(n)$, though other choices are theoretically valid.

I.1. **Generic Finite Sample Bounds.** In this section we derive finite sample bounds based on Assumption 4 below. This assumption provides sufficient conditions that are implied by a variety of settings including generalized linear models.

Assumption C4 (M-Estimation Conditions)**.** *Let $\{(Y_{ui}, X_{ui}, u \in \mathcal{U}), i = 1, \ldots, n\}$ be n i.i.d. observations of the model (I.59) and let $T_u = \text{support}(\theta_u)$ where $|T_u| \leqslant s$, $u \in \mathcal{U}$. With probability $1 - \Delta_n$ we have that for all $u \in \mathcal{U}$ there are weights $w_u = w_u(Y_u, X_u)$ and C_{un} such that:*

(a) $|\mathbb{E}_n[\partial_\theta M_u(Y_u, X_u, \theta_u) - \partial_\theta M_u(Y_u, X_u, \theta_u, a_u)]'\delta| \leqslant C_{un}\|\sqrt{w_u}X_u'\delta\|_{\mathbb{P}_n,2}$
(b) $\ell\widehat{\Psi}_{u0} \leqslant \widehat{\Psi}_u \leqslant L\widehat{\Psi}_{u0}$ *for* $\ell > 1/c$, *and let* $\tilde{c} = \frac{Lc+1}{\ell c-1}\sup_{u\in\mathcal{U}}\|\widehat{\Psi}_{u0}\|_\infty\|\widehat{\Psi}_{u0}^{-1}\|_\infty$;
(c) *for all* $\delta \in A_u$ *there is* $\bar{q}_{A_u} > 0$ *such that*

$$\begin{aligned}\mathbb{E}_n[M_u(Y_u, X_u, \theta_u + \delta)] - \mathbb{E}_n[M_u(Y_u, X_u, \theta_u)] - \mathbb{E}_n[\partial_\theta M_u(Y_u, X_u, \theta_u)]'\delta + 2C_{un}\|\sqrt{w_u}X_u'\delta\|_{\mathbb{P}_n,2} \\ \geqslant \left\{\|\sqrt{w_u}X_u'\delta\|_{\mathbb{P}_n,2}^2\right\} \wedge \left\{\bar{q}_{A_u}\|\sqrt{w_u}X_u'\delta\|_{\mathbb{P}_n,2}\right\}.\end{aligned}$$

In many applications we take the weights to be $w_u = w_u(X_u) = 1$ but we allow for more general weights. Assumption 4(a) bounds to the impact of estimating the nuisance functions uniformly over $u \in \mathcal{U}$. In the setting with s-sparse estimands, we typically $C_{un} \lesssim \{n^{-1}s\log(pn)\}^{1/2}$. The loadings

$\widehat{\Psi}_u$ are assumed larger (but not too much larger) than the ideal choice $\widehat{\Psi}_{u0}$ defined in (I.62). This is formalized in Assumption 4(b). Assumption 4(c) is an identification condition that will be imposed for particular choices of A_u and q_{A_u}. It relates to conditions in the literature derived for the case of a singleton $\mathcal{U}$ and no nuisance functions, see the restricted strong convexity[3] used in [55] and the non-linear impact coefficients used in [10] and [17].

The following results establish rates of convergence for the ℓ_1-penalized solution with estimated nuisance functions (I.60), sparsity bounds and rates of convergence for the post-selection refitted estimator (I.61). They are based on restricted eigenvalue type conditions and sparse eigenvalue conditions. The restricted set is defined as $\Delta_{2\mathbf{c},u} = \{\delta : \|\delta_{T_u}\|_1 \leqslant 2\tilde{\mathbf{c}}\|\delta_{T_u}\|_1\}$ and the restricted eigenvalue is defined as $\bar{\kappa}_{u,2\mathbf{c}} = \inf_{\delta\in\Delta_{2\mathbf{c},u}} \|\sqrt{w_u}X_u'\delta\|_{\mathbb{P}_n,2}/\|\delta_{T_u}\|$ for $u\in\mathcal{U}$. In the results for sparsity and post-selection refitted models, the minimum and maximum sparse eigenvalues, respectively

$$\phi_{\min}(m,u) = \min_{1\leqslant\|\delta\|_0\leqslant m} \frac{\|\sqrt{w_u}X_u'\delta\|^2_{\mathbb{P}_n,2}}{\|\delta\|^2} \quad\text{and}\quad \phi_{\max}(m,u) = \max_{1\leqslant\|\delta\|_0\leqslant m} \frac{\|\sqrt{w_u}X_u'\delta\|^2_{\mathbb{P}_n,2}}{\|\delta\|^2},$$

are also relevant quantities to characterize the behavior of the estimators.

Lemma 19. *Suppose that Assumption 4 holds with* $\delta \in A_u = \{\delta : \|\delta_{T_u}\|_1 \leqslant 2\tilde{\mathbf{c}}\|\delta_{T_u}\|_1\} \cup \{\delta : \|\delta\|_1 \leqslant \frac{6c\|\widehat{\Psi}_{u0}^{-1}\|_\infty}{\ell c-1}\frac{n}{\lambda}C_{un}\|\sqrt{w_u}X_u'\delta\|_{\mathbb{P}_n,2}\}$ *and* $\bar{q}_{A_u} > 3\left\{(L+\frac{1}{c})\|\widehat{\Psi}_{u0}\|_\infty\frac{\lambda\sqrt{s}}{n\bar{\kappa}_{u,2\tilde{\mathbf{c}}}} + 9\tilde{\mathbf{c}}C_{un}\right\}$. *Suppose that λ satisfies condition (I.63) with probability $1-\Delta_n$. Then, with probability $1-2\Delta_n$ we have uniformly over $u\in\mathcal{U}$*

$$\begin{aligned} \|\sqrt{w_u}X_u'(\widehat{\theta}_u-\theta_u)\|_{\mathbb{P}_n,2} &\leqslant 3\left\{(L+\tfrac{1}{c})\|\widehat{\Psi}_{u0}\|_\infty\tfrac{\lambda\sqrt{s}}{n\bar{\kappa}_{u,2\tilde{\mathbf{c}}}} + 9\tilde{\mathbf{c}}C_{un}\right\} \\ \|\widehat{\theta}_u-\theta_u\|_1 &\leqslant 3\left\{\tfrac{(1+2\tilde{\mathbf{c}})\sqrt{s}}{\bar{\kappa}_{u,2\tilde{\mathbf{c}}}} + \tfrac{6c\|\widehat{\Psi}_{u0}^{-1}\|_\infty}{\ell c-1}\tfrac{n}{\lambda}C_{un}\right\}3\left\{(L+\tfrac{1}{c})\|\widehat{\Psi}_{u0}\|_\infty\tfrac{\lambda\sqrt{s}}{n\bar{\kappa}_{u,2\tilde{\mathbf{c}}}} + 9\tilde{\mathbf{c}}C_{un}\right\}\end{aligned}$$

Lemma 20 (M-Estimation Sparsity)**.** *In addition to condition of Lemma 19, assume that with probability $1-\Delta_n$ for all $u\in\mathcal{U}$ and $\delta\in\mathbb{R}^p$ we have*

$$|\{\mathbb{E}_n[\partial_\theta M_u(Y_u,X_u,\widehat{\theta}_u) - \partial_\theta M_u(Y_u,X_u,\theta_u)]\}'\delta| \leqslant L_{un}\|\sqrt{w_u}X_u'\delta\|_{\mathbb{P}_n,2}.$$

Let $\mathcal{M}_u = \{m\in\mathbb{N} : m \geqslant 2\phi_{\max}(m,u)L_u^2\}$ *where* $L_u = \frac{c\|\widehat{\Psi}_{u0}^{-1}\|_\infty}{c\ell-1}\frac{n}{\lambda}\{C_{un}+L_{un}\}$. *Then with probability $1-3\Delta_n$ we have that*

$$\widehat{s}_u \leqslant \min_{m\in\mathcal{M}_u}\phi_{\max}(m,u)L_u^2 \quad \textit{for all}\ \ u\in\mathcal{U}.$$

Lemma 21. *Let $\widetilde{T}_u, u\in\mathcal{U}$, be the support used for post penalized estimator (I.61) and $\tilde{s}_u = |\widetilde{T}_u|$ its cardinality. In addition to condition of Lemma 19, suppose that Assumption 4(c) holds also for* $A_u = \{\delta : \|\delta\|_0 \leqslant \tilde{s}_u + s\}$ *with probability $1-\Delta_n$ with* $\bar{q}_{A_u} > 2\left\{\frac{\sqrt{\tilde{s}_u+s_u}\|\mathbb{E}_n[S_u]\|_\infty}{\sqrt{\phi_{\min}(\tilde{s}_u+s_u,u)}} + 3C_{un}\right\}$ *and* $\bar{q}_{A_u} > 2\{\mathbb{E}_n[M_u(Y_u,X_u,\tilde{\theta}_u)] - \mathbb{E}_n[M_u(Y_u,X_u,\theta_u)]\}_+^{1/2}$. *Then, we have uniformly over $u\in\mathcal{U}$*

$$\|\sqrt{w_u}X_u'(\tilde{\theta}_u-\theta_u)\|_{\mathbb{P}_n,2} \leqslant \{\mathbb{E}_n[M_u(Y_u,X_u,\tilde{\theta}_u)] - \mathbb{E}_n[M_u(Y_u,X_u,\theta_u)]\}_+^{1/2} + \frac{\sqrt{\tilde{s}_u+s_u}\|\mathbb{E}_n[S_u]\|_\infty}{\sqrt{\phi_{\min}(\tilde{s}_u+s_u,u)}} + 3C_{un}.$$

In Lemma 21, if $\widetilde{T}_u = \text{support}(\widehat{\theta}_u)$, we have that

$$\mathbb{E}_n[M_u(Y_u,X_u,\tilde{\theta}_u)] - \mathbb{E}_n[M_u(Y_u,X_u,\theta_u)] \leqslant \mathbb{E}_n[M_u(Y_u,X_u,\widehat{\theta}_u)] - \mathbb{E}_n[M_u(Y_u,X_u,\theta_u)] \leqslant \lambda C'\|\widehat{\theta}_u-\theta_u\|_1$$

and $\sup_{u\in\mathcal{U}}\|\mathbb{E}_n[S_u]\|_\infty \leqslant C'\lambda$ with high probability where $C' \leqslant L\sup_{u\in\mathcal{U}}\|\widehat{\Psi}_{u0}\|_\infty$.

[3]Assumption 4 (a) and (c) could have been stated with $\{C_{un}/\sqrt{s}\}\|\delta\|_1$ instead of $C_{un}\|\sqrt{w_u}X_u'\delta\|_{\mathbb{P}_n,2}$.

These results generalize important results of the ℓ_1-penalized estimators to the case of functional response data and estimated of nuisance functions.

A key assumption in Lemmas 19-21 is that the choice of λ satisfies (I.63). We provide next a set of simple generic conditions that will imply the validity of the proposed choice. These generic conditions can be verified in many applications of interest.

Condition WL. *For each $u \in \mathcal{U}$, let $S_u = \partial_\theta M_u(Y_u, X_u, \theta_u, a_u)$, and suppose that:*
(i) $\sup_{u\in\mathcal{U}} \max_{k\leqslant p} \{\mathrm{E}[|S_{uk}|^3]\}^{1/3}/\{\mathrm{E}[|S_{uk}|^2]\}^{1/2}\Phi^{-1}(1-\gamma/\{2pN_n\}) \leqslant \delta_n n^{1/6}$, for all $u \in \mathcal{U}$, $k \in [p]$; (ii) $N_n \geqslant N(\epsilon, \mathcal{U}, d_{\mathcal{U}})$, where ϵ is such that with probability $1-\Delta_n$:
$\sup_{d_{\mathcal{U}}(u,u')\leqslant\epsilon} \frac{\|\mathbb{E}_n[S_u - S_{u'}]\|_\infty}{\mathrm{E}[|S_{uk}|^2]^{1/2}} \leqslant \delta_n n^{-\frac{1}{2}}$, *and* $\sup_{d_{\mathcal{U}}(u,u')\leqslant\epsilon} \max_{k\leqslant p} \frac{|\mathrm{E}[S^2_{uk} - S^2_{u'k}]| + |(\mathbb{E}_n - \mathrm{E})[S^2_{uk}]|}{\mathrm{E}[|S_{uk}|^2]} \leqslant \delta_n$.

The following technical lemma justifies the choice of penalty level λ. It is based on self-normalized moderate deviation theory.

Lemma 22 (Choice of λ)**.** *Suppose Condition WL holds, let $c' > c > 1$ be constants, $\gamma \in [1/n, 1/\log n]$, and $\lambda = c'\sqrt{n}\Phi^{-1}(1-\gamma/\{2pN_n\})$. Then for $n \geqslant n_0$ large enough depending only on Condition WL,*

$$\mathrm{P}\left(\lambda/n \geqslant c \sup_{u\in\mathcal{U}} \|\widehat{\Psi}^{-1}_{u0}\mathbb{E}_n[\partial_\theta M_u(Y_u, X_u, \theta_u, a_u)]\|_\infty\right) \geqslant 1 - \gamma - o(\gamma) - \Delta_n.$$

We note that Condition WL(iii) contains high level conditions. See [] for examples that satisfy these conditions. The following corollary summarizes these results for many applications of interest in well behaved designs.

Corollary 5 (Rates under Simple Conditions)**.** *Suppose that with probability $1 - o(1)$ we have that $C_{un} \vee L_{un} \leqslant C\{n^{-1}s\log(pn)\}^{1/2}$, $(Lc+1)/(\ell c - 1) \leqslant C$, $w_u = 1$, and Condition WL holds with $\log N_n \leqslant C\log(pn)$. Further suppose that with probability $1 - o(1)$ the sparse minimal and maximal eigenvalues are well behaved, $c \leqslant \phi_{\min}(s\ell_n, u) \leqslant \phi_{\max}(s\ell_n, u) \leqslant C$ for some $\ell_n \to \infty$ uniformly over $u \in \mathcal{U}$. Then with probability $1 - o(1)$ we have*

$$\sup_{u\in\mathcal{U}} \|X_u'(\widehat{\theta}_u - \theta_u)\|_{\mathbb{P}_n,2} \lesssim \sqrt{\frac{s\log(pn)}{n}}, \quad \sup_{u\in\mathcal{U}} \|\widehat{\theta}_u - \theta_u\|_1 \lesssim \sqrt{\frac{s^2\log(pn)}{n}}, \quad \textit{and} \quad \sup_{u\in\mathcal{U}} \|\widehat{\theta}_u\|_0 \lesssim s.$$

Moreover, if $\widetilde{T}_u = \mathrm{support}(\widehat{\theta}_u)$, we have that

$$\sup_{u\in\mathcal{U}} \|X_u'(\tilde{\theta}_u - \theta_u)\|_{\mathbb{P}_n,2} \lesssim \sqrt{\frac{s\log(pn)}{n}}$$

Appendix J. Bounds on Covering entropy

Let $(W_i)_{i=1}^n$ be a sequence of independent copies of a random element W taking values in a measurable space $(\mathcal{W}, \mathcal{A}_{\mathcal{W}})$ according to a probability law P. Let $\mathcal{F}$ be a set of suitably measurable functions $f\colon \mathcal{W} \to \mathbb{R}$, equipped with a measurable envelope $F\colon \mathcal{W} \to \mathbb{R}$.

Lemma 23 (Algebra for Covering Entropies)**.** *Work with the setup above.*
(1) Let $\mathcal{F}$ be a VC subgraph class with a finite VC index k or any other class whose entropy is bounded

above by that of such a VC subgraph class, then the uniform entropy numbers of $\mathcal{F}$ obey

$$\sup_Q \log N(\epsilon\|F\|_{Q,2}, \mathcal{F}, \|\cdot\|_{Q,2}) \lesssim 1 + k\log(1/\epsilon) \vee 0$$

(2) For any measurable classes of functions $\mathcal{F}$ and $\mathcal{F}'$ mapping $\mathcal{W}$ to $\mathbb{R}$,

$$\log N(\epsilon\|F+F'\|_{Q,2}, \mathcal{F}+\mathcal{F}', \|\cdot\|_{Q,2}) \leqslant \log N\left(\tfrac{\epsilon}{2}\|F\|_{Q,2}, \mathcal{F}, \|\cdot\|_{Q,2}\right) + \log N\left(\tfrac{\epsilon}{2}\|F'\|_{Q,2}, \mathcal{F}', \|\cdot\|_{Q,2}\right),$$

$$\log N(\epsilon\|F\cdot F'\|_{Q,2}, \mathcal{F}\cdot\mathcal{F}', \|\cdot\|_{Q,2}) \leqslant \log N\left(\tfrac{\epsilon}{2}\|F\|_{Q,2}, \mathcal{F}, \|\cdot\|_{Q,2}\right) + \log N\left(\tfrac{\epsilon}{2}\|F'\|_{Q,2}, \mathcal{F}', \|\cdot\|_{Q,2}\right),$$

$$N(\epsilon\|F\vee F'\|_{Q,2}, \mathcal{F}\cup\mathcal{F}', \|\cdot\|_{Q,2}) \leqslant N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, \|\cdot\|_{Q,2}\right) + N\left(\epsilon\|F'\|_{Q,2}, \mathcal{F}', \|\cdot\|_{Q,2}\right).$$

(3) For any measurable class of functions $\mathcal{F}$ and a fixed function f mapping $\mathcal{W}$ to $\mathbb{R}$,

$$\log\sup_Q N(\epsilon\||f|\cdot F\|_{Q,2}, f\cdot\mathcal{F}, \|\cdot\|_{Q,2}) \leqslant \log\sup_Q N\left(\epsilon/2\|F\|_{Q,2}, \mathcal{F}, \|\cdot\|_{Q,2}\right)$$

(4) Given measurable classes $\mathcal{F}_j$ and envelopes F_j, $j=1,\ldots,k$, mapping $\mathcal{W}$ to $\mathbb{R}$, a function $\phi\colon \mathbb{R}^k\to\mathbb{R}$ such that for $f_j, g_j \in \mathcal{F}_j$, $|\phi(f_1,\ldots,f_k) - \phi(g_1,\ldots,g_k)| \leqslant \sum_{j=1}^k L_j(x)|f_j(x)-g_j(x)|$, $L_j(x)\geqslant 0$, and fixed functions $\bar f_j \in \mathcal{F}_j$, the class of functions $\mathcal{L} = \{\phi(f_1,\ldots,f_k) - \phi(\bar f_1,\ldots,\bar f_k)\colon f_j\in\mathcal{F}_j, j=1,\ldots,k\}$ satisfies

$$\log\sup_Q N\left(\epsilon\Big\|\sum_{j=1}^k L_jF_j\Big\|_{Q,2}, \mathcal{L}, \|\cdot\|_{Q,2}\right) \leqslant \sum_{j=1}^k \log\sup_Q N\left(\tfrac{\epsilon}{k}\|F_j\|_{Q,2}, \mathcal{F}_j, \|\cdot\|_{Q,2}\right).$$

Proof. See Lemma L.1 in [16]. ■

Lemma 24 (Covering Entropy for Classes obtained as Conditional Expectations)**.** *Let $\mathcal{F}$ denote a class of measurable functions $f\colon\mathcal{W}\times\mathcal{Y}\to\mathbb{R}$ with a measurable envelope F. For a given $f\in\mathcal{F}$, let $\bar f\colon\mathcal{W}\to\mathbb{R}$ be the function $\bar f(w) := \int f(w,y)d\mu_w(y)$ where μ_w is a regular conditional probability distribution over $y\in\mathcal{Y}$ conditional on $w\in\mathcal{W}$. Set $\bar{\mathcal{F}} = \{\bar f\colon f\in\mathcal{F}\}$ and let $\bar F(w) := \int F(w,y)d\mu_w(y)$ be an envelope for $\bar{\mathcal{F}}$. Then, for $r,s\geqslant 1$,*

$$\log\sup_Q N(\epsilon\|\bar F\|_{Q,r}, \bar{\mathcal{F}}, \|\cdot\|_{Q,r}) \leqslant \log\sup_{\widetilde Q} N((\epsilon/4)^r\|F\|_{\widetilde Q,s}, \mathcal{F}, \|\cdot\|_{\widetilde Q,s}),$$

where Q belongs to the set of finitely-discrete probability measures over $\mathcal{W}$ such that $0<\|\bar F\|_{Q,r}<\infty$, and $\widetilde Q$ belongs to the set of finitely-discrete probability measures over $\mathcal{W}\times\mathcal{Y}$ such that $0<\|F\|_{\widetilde Q,s}<\infty$. In particular, for every $\epsilon>0$ and any $k\geqslant 1$,

$$\log\sup_Q N(\epsilon, \bar{\mathcal{F}}, \|\cdot\|_{Q,k}) \leqslant \log\sup_{\widetilde Q} N(\epsilon/2, \mathcal{F}, \|\cdot\|_{\widetilde Q,k}).$$

Proof. See Lemma L.2 in [16]. ■